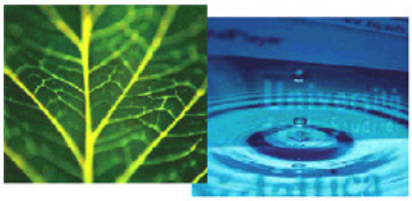
PhD Dissertation



International Doctorate School in Information and Communication Technologies

DISI - University of Trento

SENSING SOCIAL INTERACTIONS USING NON-VISUAL AND NON-AUDITORY MOBILE SOURCES, MAXIMIZING PRIVACY AND MINIMIZING OBTRUSIVENESS

Aleksandar Matic

Advisors:
Dr. Venet Osmani
Dr. Oscar Mayora
Prof. Imrich Chlamtac

May 2012

Abstract

Social interaction is one of the basic components of human life which impacts thoughts, emotions, decisions, and the overall wellbeing of individuals. In this regard, monitoring social activity constitutes an important factor for a number of disciplines, particularly the ones related to social and health sciences. Sensor-based social interaction data collection has been seen as a groundbreaking tool which has the potential to overcome the drawbacks of traditional self-reporting methods and to revolutionize social behavior analysis. However, monitoring social interactions typically implies a trade-off between the quality of collected data and the levels of unobtrusiveness and of privacy respecting, aspects which can affect spontaneity in subjects' behavior. Despite the substantial research in the area of automatic recording of social interactions, the existing solutions remain limited: they either capture audio/video data which may raise privacy concerns in monitored subjects and may restrict the application to very specific areas, or provide low accuracy in detecting social interactions that occur on small spatio-temporal scale.

The objective of this thesis is to provide and evaluate a solution for mobile monitoring of face-to-face social interactions, which maximizes privacy and minimizes obtrusiveness. In order to reliably detect social interactions that occur on small spatio-temporal scale, the proposed solution infers two types of information, namely spatial settings between subjects and their speech activity status. The challenge was to select appropriate sources that do not restrict application scenarios only to certain areas and do not capture privacy sensitive data, which are the drawbacks of video/audio systems. The second stage was to interpret the data acquired from non-visual and non-auditory sources and to model social interactions on small space- and time- scales. The work in this thesis assesses the reliability of the proposed approach in several scenarios, demonstrating the accuracy of approximately 90% in detecting the occurrence of face-to-face social interactions.

The feasibility of using the proposed approach for social interaction data collection is further evaluated with respect to the study of social psychology, which

serves as the guideline for extracting the relevant features of social interactions. The evaluation has demonstrated the possibility to extract various nonverbal behavioral cues related to spatial organization between individuals and their vocal behavior in social interactions. By modeling social context using the extracted features, it is possible to achieve the accuracy of 81% in the automatic classification between formal versus informal social interactions. In addition, the proposed approach was applied to gather daily patterns of social activity for investigating their correlation with the mood changes in individuals, which has been explored so far only using the traditional self-reporting methods. The findings are consistent with previous studies thus indicating the possibility to use the proposed method of collecting social interaction data for investigating psychological effects of social activities.

Keywords:

monitoring social interactions, wearable computing, nonverbal behavior analysis, formal and informal social context, positive and negative affect

DEDICATED TO MY FATHER

Contents

Chapter 1

1. Introduction

"Man is by nature a social animal; an individual who is unsocial naturally and not accidentally is either beneath our notice or more than human. Society is something that precedes the individual. Anyone who either cannot lead the common life or is so self-sufficient as not to need to, and therefore does not partake of society, is either a beast or a god."

Aristotle (384 BC – 322 BC)

Social behavior has been the subject of vigorous observations, debates, and analysis of thinkers and philosophers even 2,500 years ago, who attempted to describe and classify such a basic and core human phenomena. Throughout the history the social interaction was the aspect of humanity analyzed by intellectuals from nearly all disciplines, from philosophy and psychology to arts and medicine. Despite the fact that the efforts and keen interest of humanists to understand social behavior dates back to the times of ancient civilizations, it is significant to notice that first incidences of scientific data collection on human interactions took place in the beginning of the 20th century [1][2][3]. At that time an emerging discipline of social psychology aimed to understand individual and group behavior in social context relying on the data collected through surveys or by engaging a human observer who was taking notes about social interactions within monitored groups. Pioneering experimental findings already had a multidisciplinary impact on human resource movements [2], the study of children [3], and leadership styles [4], thus demonstrating the importance of social interaction data collection. Nowadays, almost a century later, the same methods for analyzing social behavior are still prevalent in social and health sciences although they exhibit a number of drawbacks. Periodical surveys, diaries and similar self-reporting methods suffer from memory dependence, recall bias and a high end user effort for continuous long-term monitoring [5][6]. Moreover they correspond poorly to commu-

nication patterns as recorded by independent observers [7]. Albeit being a more reliable method, relying on a human observer to record social interactions in groups is inefficient particularly if the size of the group is large, the interactions occur in various physical locations, or the study requires longitudinal data collection [8].

As an alternative to annotating social interactions by hand, automatic data collection methods such as tracking mobile phone calls, SMS messages, email exchange or activity in online social networks, emerged as a result of a high penetration of electronically mediated communications in the last fifteen years. Nevertheless, an electronic communication lacks the richness of social signals transmitted during a face-to-face interaction mediated by physical proximity [9] which is still considered the most important form of communication [10]. Furthermore, establishing social behavior solely through electronically mediated interactions has been seen as its oversimplification and, in many cases, to be incorrect [5]. However, in addition to the possibility of tracking electronically mediated interactions, the last decades of technological advancements have brought also automatic tools for monitoring face-to-face social interactions. The advent of sensor based instruments for recording social activity of individuals is considered to be a critical point in the evolution of social behavior analysis, exhibiting the potential to overcome the limitations of self reporting and observational methods [5]. Buchanan [11] envisioned that sensors will transform social sciences as much as microscopes transformed medicine in the 18th and the 19th century. Undoubtedly, pervasive computing paradigms already afforded the new findings on social interaction phenomena by providing an insight into domains that are difficult or impossible to be recorded by hand-annotating methods, such as: interpersonal distances can be detected with 1mm precision; hand gestures can be analyzed 10 times per second; vocal behavior can be extracted with the resolution of 10-50ms; and body orientation can be recognized with the accuracy of 1°. However, acquiring high quality social interaction data typically remains within confines of experimental conditions.

1.1. Research question

It is interesting to note that the current systems for automatic sensing of face-to-face social interactions have relied on the same senses as human observers – the

visual and auditory, thus capturing video and/or audio data. However, the use of microphones and cameras can negatively affect the subjects' perception of privacy; besides, video systems constrain the movements of monitored subjects into the areas covered with machine vision systems. Therefore, when monitoring social interactions and, in general, human behavior there is a trade-off between the quality of collected data and ecological validity in the study. In this regard, enabling natural experimental settings depends on the level of obtrusiveness, respecting subjects' privacy and spatial restrictions. More invasive methods typically provide richer information while being prone to affecting the natural behavior in subjects and consequently the reliability of measurements. Therefore, the work in this thesis answers the following questions:

What is the extent of trade-off between the quality of obtained data and provisioning of a mobile solution for monitoring face-to-face social interactions that minimizes obtrusiveness and maximizes privacy? Can face-to-face social interactions be reliably detected without using visual and auditory sources?

In other words – while microscope restricts experiments within laboratory, do we have a mobile magnifying glass to conduct reliable field research?

1.2. Objective

The objective of the work in this thesis is to allow continuous and reliable social interaction data collection by interpreting information acquired from non-visual and non-auditory mobile sources while maximizing privacy and minimizing obtrusiveness. The focus is placed on social interactions that occur on small spatio-temporal scale i.e. co-located face-to-face conversations, and the rest of the thesis refers only to that form of social interaction. By relying on non-visual and non-auditory sources for monitoring social interactions, this work takes on the challenges of interpreting noisy data in order to maximize privacy, minimize obtrusiveness and minimize spatial limitations of experimental settings. However, despite acquiring rudimentary evidences, the proposed solution reliably detects the occurrence of social interaction and provides valuable data for further social interaction analysis.

1.3. Overview of the proposed approach

One can estimate whether two persons are having a face-to-face conversation by simply observing them from a relatively long, unobtrusive, distance and judging solely by the mutual position of their bodies. In order to ascertain if they are talking or just facing each other but not interacting, it is necessary to obtain the evidence about speech activity. Yet, approaching the earshot of monitored subjects may raise privacy concerns and consequently affect their natural behavior. Witnessing speech activity but not affecting the perception of privacy in subjects and observing them but not being obtrusive is the strategy that would lead towards capturing greater mundane realism regarding interaction data collection. This principle was followed to develop the approach proposed in this thesis, which is intended for continuous monitoring of face-to-face social interactions, while not using visual or auditory evidences.

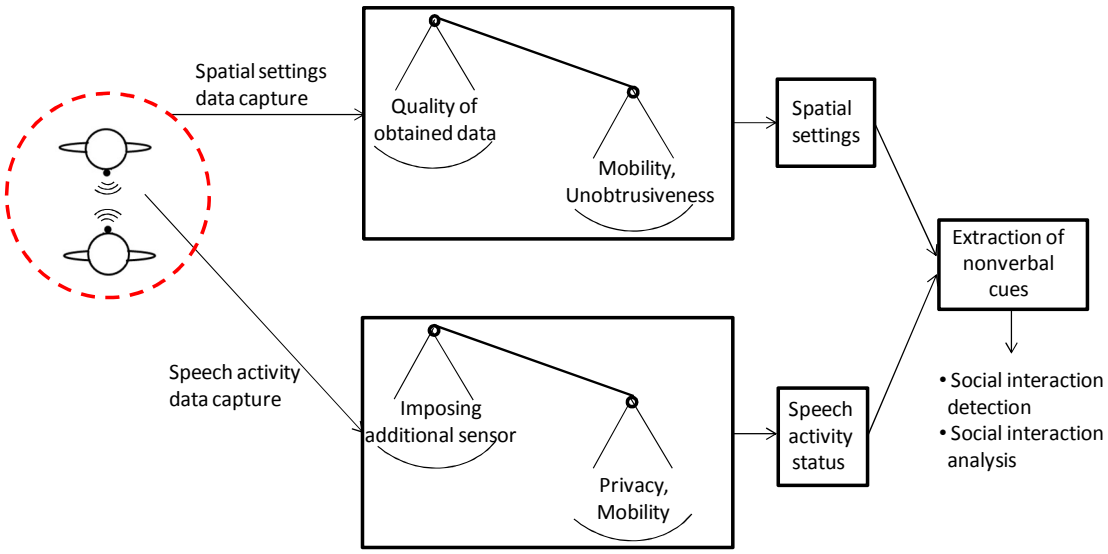


Figure 1.1 Concept of the proposed system

The first task is to infer spatial settings between subjects, described by parameters of interpersonal distance and relative body orientation (Figure 1.1). Groh et al. [12] demonstrated that these two parameters provide sufficient evidence to detect the occurrence of social interaction, however using a highly precise camera-beacon system with the accuracy of <1mm and <1°. The use of the mounted camera system was avoided in order to follow individuals continuously, regardless of their location, and not to capture video, which may contain sensitive information. Instead, the method for recognizing spatial settings between subjects, proposed in this thesis, relies on sensing

capabilities available in one of the most familiar and widely used wearable devices, the mobile phone. Being a device that is not dedicated to inferring face-to-face social interactions, the mobile phone does not provide interpersonal distances and body orientations directly such as purpose-designed camera system, rather requiring a complex interpretation of noisy data obtained from available embedded sensors. Thus, such an approach trades-off the quality of acquired information for allowing mobile and minimally obtrusive solution. However, the work in this thesis demonstrates that spatial settings parameters can be extracted by using mobile phone sensing mechanisms with a sufficiently high precision to indicate social encounters.

Since solely spatial settings do not always provide enough evidence for inferring the occurrence of social interaction [13] (for instance, the situation depicted in Figure 1.1 may also correspond to two subjects sitting across from each other in the office and not engaging in an interaction), the second task is acquiring the knowledge about speech activity of co-located subjects. In order to prevent a negative impact on the perception of privacy in monitored individuals, the approach proposed in this thesis is based on identifying another manifestation of speech different than voice, namely the vibration of vocal chords. The method relies on an external off-the-shelf accelerometer intended to infer speech activity by detecting vibrations at the chest level that are generated by vocal chords during phonation. Although a microphone embedded in the mobile phone could be used for speech detection (such as in [13]), the proposed system involves an additional sensor due to the fact that activating microphone may raise privacy concerns with subjects despite privacy sensitive techniques, thus affecting their natural behavior. Moreover, nearby conversations, in which the monitored subjects do not participate, can be unintentionally picked up by the microphone. The accelerometer-based approach does not require capturing sensitive information but, on the other hand, faces the challenges of interpreting noisy data acquired from a source, which is not dedicated for speech activity detection. In addition, wearing a sensor at the chest level may be perceived as obtrusive and consequently it may stigmatize monitored subjects. This issue, while currently a concern, is expected to be mitigated, since accelerometers are increasingly becoming widely adopted both in research and everyday life. The shape and size of already accepted commercial accelerometer-based solutions can suit also the speech recognition purpose (such as Fitbit

[14] – an accelerometer device for tracking wellbeing aspects of individuals' behavior), while the chest area is convenient for attaching a sensor with an elastic band (similarly to attaching respiratory or cardio sensors) minimizing the interference with typical daily routines. Therefore, relying on an accelerometer as an alternative to the use of microphone was a compromise for preventing privacy concerns in subjects while providing a mobile solution for continuous monitoring of speech activity.

This work provides an approach for automatic monitoring of co-located social interactions, tackling issues from selecting sensors to interpreting the obtained data. The proposed solution infers speech activity and establishes spatial settings between them, while respecting subjects' privacy and minimizing obtrusiveness. Referring to social psychology literature, the system is capable to extract meaningful social interaction features, which is demonstrated through two studies on 1) identification of social context, perceived by subjects as formal or informal, and 2) inferring patterns of social activity that provoke similar responses in individuals' mood.

1.4. Background

"*To observe is to disturb*."
Werner Heisenberg (1901 – 1976)

Capturing spontaneous social interactions that occur in natural conditions pertains to recording people as they freely go about their lives [13]. The ultimate goal is to develop a method with the highest precision in collecting social interaction data which is fully privacy respecting and invisible from users' perspective, while not restricting the application to a limited number of scenarios. However, in practice there is typically a trade-off between these aspects – the more privacy respecting and unobtrusive the approach is, the more limited possibilities of acquiring social interaction data are [15].

1.4.1 Obtrusiveness

Although defined in quantum mechanics, Werner Heisenberg's uncertainty principle, which rejects the notion of a passive observer, may be applied also in the domain of social behavior analysis, regardless of conducting manual or automatic

method for interaction data collection. Neither hand-annotating methods nor the current sensor-based systems, including the work in this thesis, enable monitoring of social interactions without disturbing subjects. The goal of the method for collecting social interaction data is to extract the most information out of the least obtrusive sources; however this is a challenging problem since noninvasive methods typically result in the output that is difficult to be processed effectively and vice versa, invasive methods provide more detailed information that is easier to process but tend to change the behavior of monitored subjects [15].

1.4.2 Privacy

In addition to physical obtrusiveness, monitoring of human behavior is often closely linked to disturbing one's privacy. Privacy issues relate to an array of ethical norms that need to be addressed. All subjects in the study should always know that they are being monitored, moreover they must have the right to authorize the use and the diffusion of the collected data [15]. If monitoring involves audio or video archives, they can be partially or totally deleted by subjects while recording uninvolved parties without their consent is considered unethical and illegal [13]. However, despite addressing all the ethical norms, people are prone to change their behavior if they have concerns about the way of monitoring, which negatively affects the reliability of the collected data. In the case of old methods, such concerns can be raised due to a human observer, while for sensor-based approaches the presence of audio/video data analysis becomes an issue to consider. When automatically recording social interactions, protecting privacy often implies discarding sociologically useful information [13] which is not always an acceptable compromise. However, even though privacy sensitive recording techniques are applied, the fact that a microphone or a camera is activated may still raise concerns. This often depends on the technical education and cultural background of monitored subjects, which can affect their perception of privacy [16][17].

Regardless whether applying observational, self-reporting or sensor-based method for collecting interaction data, when sensing social interactions and, in general, human behavior, important issues are the levels of subjects' privacy and of obtrusiveness of the method [15]. The problem illuminates a well-known trade-off between

the spectrum and quality of collected data and enabling natural conditions, where typically the solution reflects the trade-off.

1.5. Motivation

Monitoring of social interactions represents an important aspect for social behavior analysis, a domain which has a wide-reaching, multi-disciplinary impact. These disciplines range from medicine where quantitative evaluation of social activity represents a tool in coaching and diagnosis [12], to economics where social relationships are used to model both micro- and macroeconomic phenomena [18], to anthropology, which analyzes differences in social behavior across different cultures [19], to epidemiology which examines interpersonal contacts as the main cause behind spreading of an epidemic [20]. However, the work in this thesis was greatly motivated by the research initiatives on understanding social behavior in two disciplines, namely social psychology and ubiquitous computing.

Social psychology is the scientific discipline which studies how individuals' thoughts, feelings, and behaviors are influenced by the presence of other people [21]. Since exploring the phenomenon of social facilitation in 1898 [1] considered to be the first published experimental study in this area, social psychologists have examined various aspects related to interaction dynamics, formation, structure and performance of small groups (including status, norms, roles, productivity and decision making) [22]. In addition to investigating dynamics of social behavior, the focus of social psychology is also on nonverbal behavior in human interactions, shown to be crucial in inference of emotions, attitudes, relationships and traits of individuals [9][15][22][23][24]. However, typically small amounts of manually annotated interaction data limit the research in interaction dynamics and nonverbal cues analysis. By enabling collection of larger amounts of interaction data and, at the same time providing more precise insight into domains of social behavior, technological solutions have potential to advance the knowledge in social psychology. Referring to the previous literature helps identifying nonverbal cues that are meaningful for interpreting behavior [8] thus setting goals for the research in ubiquitous computing of what needs to be extracted for the analysis.

Mark Weiser, who coined the term "ubiquitous computing" in 1988, envisioned physical environments and everyday objects with computational and networking facilities, supporting people in everyday life. According to his vision, researchers have been working on sensor-based approaches to identify subjects, recognize their activities and context, offering to users a wide-range of services that are adaptive and responsive to their needs, habits or the environmental factors. Nevertheless, sensing technologies are always moving towards further miniaturization and increased computational power, constantly challenging researchers to creatively exploit new sensing potentials while dealing with sensor uncertainty and noise. One outcome of the research in ubiquitous computing is human behavior analysis with an important focus on social interactions. Along this line, reality mining, the approach of analyzing human behavior and social patterns at large scale [25], has been identified as one of the 10 technologies with a potential to change the world [26]. More generally, social signal processing is an emerging technological domain that aims to provide computers with the ability to sense and understand social signals [27]. However, unobtrusiveness and privacy respecting are the aspects of sensor-based monitoring which are important both for users and also for researchers who aim to decrease the observation effect on the monitored subjects. The goal is to attain the highest possible spontaneity of the behavior in the subjects, which becomes a very challenging and a well-known problem regarding data collection.

1.6. Contributions and thesis structure

The main contributions of this thesis are:

- proposing and evaluating a new mobile-based approach for social interaction data collection, which does not capture privacy-sensitive data and does not interfere with most of typical daily activities of individuals;
- the method for inferring spatial settings and speech activity of subjects by interpreting data obtained from non-visual and non-auditory sources;
- assessing possibilities of nonverbal cues extraction and demonstrating the high predictive power of several cues for detecting the occurrence of social

interactions on small spatio-temporal scale and for classifying the social context;

- initial validation of the proposed approach to analyze social activity as it correlates with mood changes

This thesis is organized as follows.

Chapter 2 reviews the state of the art in the field of automatic sensing of social interactions. Relevant studies are divided according to the utilized sensors which can be wearable or part of external infrastructure.

Chapter 3 presents and evaluates the approach for using mobile phones to estimate interpersonal distances and relative body orientations. The evaluation is conducted across multiple environments and using diverse phone models in order to assess the accuracy of the method and to identify factors which can cause accuracy degradation.

Chapter 4 describes the accelerometer-based approach for speech activity detection. The approach was assessed during mild and intense activities and in several vehicles whose engines may cause false positives, including car, train, bus, airplane and elevator.

Chapter 5 provides the evaluation of the detection of face-to-face social interactions using the two proposed sensing modalities, namely spatial settings detection and speech activity status recognition. The experiments were conducted in both controlled and continuous real-life settings, indicating performances of each modality in isolation and of their fusion.

Chapter 6 examines which features of social interactions are relevant for distinguishing between formal and informal context according to the theory of social psychology. Afterwards, this chapter discusses the possibilities of extracting the relevant features using the proposed solution. Finally, it presents the performance in the automatic classification between formal and informal contexts.

Chapter 7 describes the use of the proposed solution for collecting social interaction data in three experimental trials, conducted with the goal to investigate the correlations between patterns of social activities at workplace and the mood changes of workers. After each presented trial, the results are compared with the current studies reported in the social psychology literature which relied on survey-based methods.

Chapter 8 summarizes the main contributions of this thesis and indicates future directions.

Chapter 2

2. State of the art in sensing social interactions

Steady decrease in device form factor, coupled with an increase in computational capabilities, has enabled monitoring of many aspects of social behavior, from quantifying dynamics of social activity to extracting various social signals expressed during social interactions. The choice of sensors and their arrangement in experimental settings determines the level of privacy, obtrusiveness and the spectrum of interaction data that can be extracted. The use of video/audio infrastructure, wearable dedicated hardware or mobile phones provide different trade-offs between the quality of collected data and the constraints for experimental settings. It is difficult to compare methods for collecting interaction data on a general basis; however, with these issues in mind, Table 2.1 provides a relative comparison between different methods regarding concepts that are typically applied, which is reviewed in this section in more detail.

Table 2.1: Summary of methods for collecting social interaction data

Method for collecting data	**Accuracy in detecting social interaction occurrence**	**Spectrum of extracted non-verbal behavioral cues**	**Spatial limitations**	**Privacy concerns**	**Obtrusiveness**
Survey based	Low	Low	Low	Low	Medium
Human observer	High	Medium	Medium-High	High	Medium
Video/Audio Infrastructures	High	High	High	High	Low-Medium
Dedicated devices	High	Low-Medium	Low	Medium-High	High
Mobile phones	Low (medium)*	Low	Low	Low (high)*	Low
Proposed system	High	Medium	Low	Low	Low-Medium

*using audio data

2.1. Video/audio infrastructures

This section reviews the work done in monitoring social interaction by relying on the video/audio infrastructures. Infrastructure refers to the equipment installed for a specific scenario (rather than for a longitudinal study), in order to track social interactions and to extract behavioral cues for the analysis. The common equipment includes cameras and/or microphones mounted in the area of interest and arranged in a structure that suits the objective of investigation. Such systems vary in complexity, from a single camera/microphone to fully equipped smart meeting rooms capable to capture complex group interactions [15]. Extracted behavioral cues are used to model interaction management (such as turn-taking patterns and the problem of addressing), internal states (such as interest), personalities (such as dominance and extroversion), social relations (such as roles and status), which is a domain thoroughly reviewed by Gatica-Perez in [24]. In addition, automatic analysis of behavioral cues relates to the emerging area of Social Signal Processing (SSP) which aims to bridge the social intelligence gap between computers and humans i.e. to provide computers the ability to sense and understand human social signals [9][15].

Significant results have been achieved in automatic recognition of physical appearance, gesture and posture, facial and eyes behavior, vocal behavior and spatial settings (extensive surveys of this domain from the perspective of SSP, conducted by Vinciarelli et al., can be found in [15] and [27]). Automatic sensing of physical appearance was examined in a few studies, extracting the color of skin, hair and clothes [28][29][30] and there were also attempts in measuring the beauty of faces [31][32]. Gesture recognition was addressed regarding automatic understanding of sign languages [33] or controlling computer through arm movements as an alternative to the use of keyboard and mouse [34]. In addition, Cristiani et al. [35] used automated gesture analysis to detect speech activity. Posture detection was mostly directed towards activity recognition or surveillance; the review of this field can be found in [36] and [37]. A number of works addressed the recognition of facial and eyes behavior along several research areas: identifying faces in the picture, capturing facial features or the motion of eyes/facial parts, classifying extracted information into interpretative categories (for instance, regarding emotions, social signals and facial muscle actions) [15].

Vocal behavior was analyzed through detected patterns of speech and silences, voice quality (pitch, tempo and energy) and vocalization (non-linguistic such as laughing or crying, and linguistic such as detecting hesitations in speech) [15]. Detecting spatial settings was mostly realized through machine video systems finding its application in a number of studies including the inference of social relations [38], intelligent video surveillance [39] and detection of social interactions [12].

Automatic video/audio analysis of face-to-face social interactions extracts an ample spectrum of information which can provide a high scientific and technological value. Since subjects are not required to wear sensors, such systems allow monitoring which is not physically intrusive (except for the cases when it is needed to attach microphones with headsets onto the subjects). However, the use of video/audio systems typically implies moving restrictions to the monitored subjects since video analysis requires a direct line of sight between subjects and cameras whereas the audio data is usually captured from single/multiple microphones installed in the area of interest. In addition, video and/or audio data can contain privacy sensitive information which creates additional issues when monitoring social interactions.

Summarizing, video/audio analysis provides precise insight into behavioral aspects of subjects during social interactions in a non-obtrusive manner; on the other hand, such monitoring limits applications to certain areas and captures sensitive data which may affect the spontaneity of subjects (Table 2.1). Extracting social signals is not limited only to external infrastructures, however the current wearable solutions are still not commonly utilized for SSP since they do not provide suitable quality of video and audio data [12]. Therefore, wearable solutions are mostly used for recording the occurrences of social interactions and for quantifying dynamics of social activity on a long-term scale.

2.2. Wearable devices

2.2.1 Dedicated hardware

Several purpose-built wearable devices have been developed for monitoring specific aspects of social interactions or aiming to support communication among individuals.

A leader amongst wearable devices for tracking face-to-face interactions is Sociometric Badge [10][40] (Figure 2.1) which was successfully applied in a number of studies, including investigation of productivity in enterprises [41], job satisfaction [42], and individual personalities [43]. Sociometric Badge is a pendant-like hardware that is worn at the chest-level, which can detect the occurrence of face-to-face interactions with a high accuracy. It achieves this through the use of an infrared sensor (IR) that can detect another badge when in a direct line of sight up to 1m and within a cone of 30°. The device has a microphone which can record the audio data or can capture only speech features including volume, frequency, and speaking time. In addition, body movements and physical activities are recognized using the embedded accelerometer, while the proximity to non-monitored people can be sensed by Bluetooth scanning.

Figure 2.1: Sociometric badge developed by MIT Media Lab [10][40]

Sociopatterns sensing platform [44] relies on a small active Radio Frequency Identification (RFID) tag which estimates the proximity of other tags by exchanging low-power radio packets. When setting the weakest power level, packets can be detected only when subjects are facing each other within 1m, which was used to identify social interactions assuming that RFID tags are worn at the frontal side of the body. The approach was applied in modeling the propagation of diseases and for evaluating

control measures in a primary school [45], for measuring social patterns in healthcare settings [46] and social dynamics in conferences [47].

The Electronically activated recorder (EAR) is the concept of sampling ambient sounds from a wearable recorder [48], [49] with the goal to capture a set of behavioral aspects of a user from daily acoustic logs. The recognized events are mostly related to socializing, including talk, laugher and arguments among individuals. As the audio recording technologies were progressing, the EAR has undergone three generations of improvement since 1998, from an analog micro-tape recorder to a software-based solutions installed on modern mobile devices. This concept was used in various studies on investigating differences in social interaction patterns between different cultures [50], genders [51] or personalities [52], the impact of social interactions on well-being and mood disorders [53], and in exploring phenomena such as narcissism [54] or job disengagement [55].

Several projects aimed to facilitate social interactions during group events, typically by providing the information to subjects about other individuals in their close proximity which share common interests. Although the goal of stimulating face-to-face interactions is not entirely related to the work in this thesis, this line of investigation is reviewed due to the necessarily performed task prior to supporting a conversation – detecting potential social situations. One of the pioneering projects was nTag intended to improve networking by displaying items of mutual interest on tags when two subjects face each other. The badges use IR to detect face-to-face orientation and proximity between subjects, while embedded semi-passive RFID tags provide conference organizers with security information, the attendance in sessions and the number of participants in certain areas at the conference venue [56]. Along the same line, SpotMe system was designed as a wearable device that informs users of who is standing within 30 meters radius and if a person with same interests step within 10 meters. Radio frequency based communication allows for exchanging messages or electronic business cards between users. Another similar device, called IntelliBadge, relied on RF location markers installed at the points of interests thus inferring position of individuals and their proximity during events. Since the device did not have a display, public screens were used for providing useful information to attendees. SpotMe [57] and nTag (now called Intelligent Events [58]) have recently reappeared in advanced

forms providing tools for improved networking among conference attendees, logistic information such as schedule and map of a venue, and to electronically collection of surveys. A few projects aimed to support face-to-face interactions at workplace such as Hummingbird [59], a mobile RF device which alerts users when other group members are close.

For achieving a high accuracy in detecting the occurrence of face-to-face social interactions in a mobile way, the knowledge of both proximity between subjects and their speech activity status is required (such as in the case of Sociometric Badge). In order to infer speech activity status, typical approach is audio analysis which, as previously discussed, often faces ethical issues and privacy concerns. Besides, most of dedicated devices for inferring face-to-face contacts require a direct line of sight between two units which imposes a specific position on the body for their placement; therefore such approaches are prone to affect the natural behavior of the subjects since they can interfere with daily activities.

In sum, utilizing dedicated devices for tracking social interactions provides a high accuracy in detecting social interactions, allows continuous monitoring without spatial limitations in comparison to video/audio infrastructures, and in a few cases provides the possibility to extract certain social signals (such as speech features using EAR and Sociometric Badge [10]) - Table 2.1. Approaches based on the use of dedicated hardware may cause privacy concerns in monitored subjects due to the fact that achieving a higher accuracy in detecting face-to-face conversations requires activating a microphone. Moreover, dedicated devices can be perceived as obtrusive from the perspective of subjects that should wear them on visible places of the frontal part of the body.

One way to address the issue of stigmatizing subjects is to utilize the sensing capabilities available in one of the most familiar device – the mobile phone. Current work on mobile phone sensing to infer social activity is reviewed in the following.

2.2.2 Mobile phone sensing

The rapid adoption of mobile phones brings the opportunity for an unobtrusive and continuous monitoring of social interactions and, in general, individuals' behavior [5]. The challenge is how to address monitoring of specific activities relying on existing sensing technologies that are embedded in mobile phones, which is the issue not

encountered when using purpose-manufactured devices which already have dedicated sensors incorporated.

Current work on mobile phone sensing to detect social interactions has relied mostly on using Bluetooth to sense nearby mobile phones. Using Bluetooth as a proximity sensor to reconstruct social dynamics at large scale has been extensively investigated under the umbrella of reality mining initiative [60][61][5]. Since the Bluetooth communications range is in the order of ten meters, this approach provides only a coarse spatial granularity in recognizing interpersonal distances; therefore, the knowledge about proximity between individuals is used to model the dynamics of social interactions at large scale rather than detecting each single social encounter which takes place at small spatio-temporal scale. MIT Media Lab's Reality Mining project launched in 2004 with the goal of sensing complex social systems which included inferring patterns in daily user activity, relationships, socially meaningful locations, and organizational structures [60]. Along the same line, Raento et al. [62] were one of the first who proposed mobile phone data collection for large-scale context sensing. More recent algorithm for identifying social groups and inferring frequency/duration of meetings within each group was proposed by Mardenfeld et al. [63] who tested their approach on the Reality Mining dataset. In addition to modeling the patterns of person-to-person interactions, Do and Gatica-Perez [64] showed that it is possible to infer different interaction types using a probabilistic model applied on longitudinal Bluetooth data.

In order to address the limitation of Bluetooth scan to detect actual face-to-face proximity between subjects, the Virtual Compass project [65] estimates interpersonal distances using RSSI analysis of Bluetooth and Wi-Fi signals. By applying empirical propagation models, the approach achieves the median accuracy between 0.9 m and 1.9 m while also detecting position of subjects in 2D plane. However, the lack of subjects' orientation information and the lack of the knowledge of speech activity might not be sufficient for modeling the occurrence of face-to-face social interactions. One of the recent works on detecting face-to-face conversations using mobile phones, described in [13], demonstrated that interactions inferred from speech and co-location are different. The authors proposed the method of extracting audio data features using microphones from a pair of co-located mobile phones, in order to detect who was

speaking and when thus detecting face-to-face interactions. The algorithm does not capture raw audio data but a set of features which does not contain verbal information. However, the limitations of this approach include: 1) sensitivity to false positives since the conversations occurring in close proximity of the monitored subjects in which they are not involved, can be incorrectly classified, 2) activating microphone can negatively affect the perception of privacy in subjects.

To recap, mobile phones have shown to be a platform for an unobtrusive, privacy respecting and continuous sensing of proximity of subjects which was further used to quantify social interactions of individuals at large scale. Such an approach does not suffice for detecting social interactions that occur on small spatio-temporal scale. As an alternative, the use of a microphone embedded in mobile phones provides a solution for inferring real-time face-to-face conversations, with the drawbacks of causing privacy concerns and unintentional picking up of the nearby conversations in which the monitored subject are not involved.

2.3. Summary

The existing solutions for recognizing social interactions remain limited with respect to the following aspects: they either require expensive infrastructures which spatially constrain applications, involve devices that are often not available off-the-shelf, provide limited accuracy in gathering real-time data with spatial and temporal granularities, or make the use of microphone whose activation may raise privacy concerns in monitored subjects.

Chapter 3

3. Inferring interpersonal spatial settings

Social interaction is not only spoken words – in parallel a wealth of information is conveyed nonverbally through interpersonal distances, body gestures and posture, eye gaze, tone of the voice, and facial expressions [24]. Through these social signals, expressed as temporal patterns of nonverbal cues, people communicate emotions, relationships, and attitudes and also infer traits of other participants in social interactions [22][9]. One of the most significant nonverbal cues related to social and emotional closeness is the interpersonal distance [12], reflecting also other phenomena such as the cultural background and the type of personality. Besides, setting an appropriate interpersonal distance is a prerequisite for carrying out a face-to-face conversation (for instance, it is unlikely that in an ordinary interaction two people communicate standing on 5 m distance), thus the distance between subjects becomes also a meaningful evidence of the existence of social interaction. In particular, the high predictive power of the interpersonal distance when combined with the relative body orientation (both parameters detected using a camera system) was demonstrated for inferring an ongoing face-to-face interaction in [12], . Therefore, the two parameters, namely interpersonal distance and the relative body orientation (which is a dimension of posture) were selected to represent spatial settings between subjects.

The criterion for achieving continuous interaction data collection excludes the choice of using video/audio infrastructures (which limit monitoring only to certain areas), self-reporting methods (which have difficulties in capturing single occurrences of social interactions) or human observers (who typically have limited set of locations where to follow subjects). These limitations lead to the choice of a mobile wearable solution, which allows tracking subjects regardless of their location. However, the criteria of mitigating privacy concerns and minimizing obtrusiveness renders unsuitable use of currently available dedicated devices, since they typically use the microphone and may also physically interfere with subjects' daily routines. On the other hand,

mobile phones are already ubiquitous devices that have been adopted faster than any technology in human history [5]. The process of monitoring behavior conducted through mobile phones fades into the background, allowing continuous data collection with a minimal effect on the users' behavior and consequently, their social interaction patterns.

However, relying on mobile phone sensing to infer spatial settings creates unique challenges which are not encountered when dedicated monitoring devices are used or when this task is shifted to the infrastructure such as machine vision systems. These challenges, chief of which is estimation of interpersonal distances and relative body orientation, are addressed in this chapter by exploiting available mobile phone sensors. Whether the achieved accuracy is sufficient for indicating the occurrence of a face-to-face interaction is analyzed in chapter 5.

3.1. Estimating interpersonal distances

3.1.1 Interpersonal distances and social interactions

Regarding distances between people as they interact, a psychologist Robert Sommer explained that "the best way to learn the location of invisible boundaries is to keep walking until somebody complains" [66]. Yet, the most commonly used classification of interpersonal distances hold in social interactions was defined by the study of proxemics. Proxemics, the term originally coined by Hall [19], refers to the study of the communicative function of space characterized as an out-of-awareness distance-setting. Hall defined four categories of interpersonal distances, including close phase (denoted with c) and far phase (denoted with f). For North American culture, categories of interpersonal distances include the following metrics (Figure 3.1): intimate distance (c: 0 – 0.15 m, f: 0.15 – 0.45 m), personal distance (c: 0.45 – 0.76 m, f: 0.76 – 1.2 m), social distance (c: 1.2 – 2.1 m, f: 2.1 – 3.6 m) and public distance (c: 3.6 – 7.6 m, f: 7.6 and more) [67]. According to the study of proxemics, these four categories of space are typically used for the following activities: intimate for embracing, touching or whispering; personal for interactions among good friends or family members; social for interactions among acquaintances; public for public speaking.

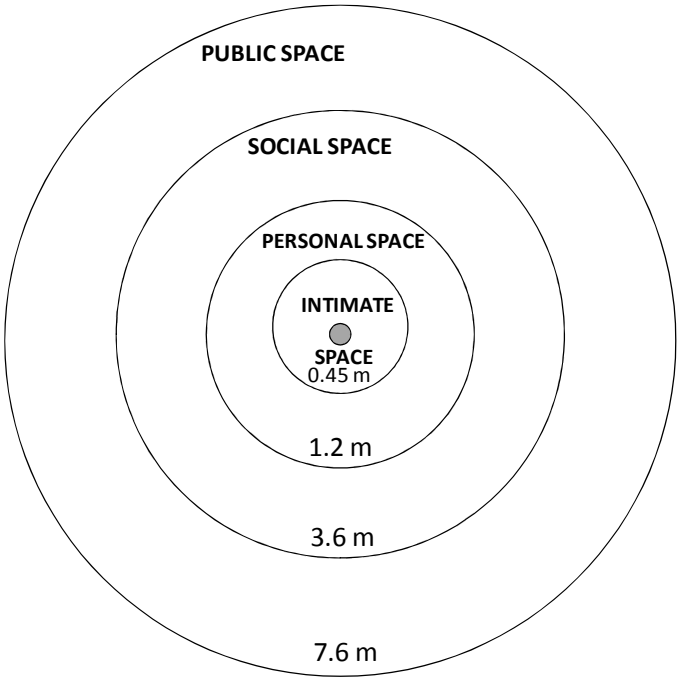


Figure 3.1: Proxemics: Categories of interpersonal distances

The recent investigation conducted by Cristani et al [38] measured interpersonal distances using computer vision techniques with the goal to investigate the fact that people tend to unconsciously organize the space around them in concentric zones corresponding to different degrees of intimacy. The results reflected the main postulates of proxemics confirming that interpersonal distances depend on the social closeness between individuals i.e. whether they are acquainted, friends or in a romantic relationship. In addition to social and emotional closeness, other parameters such as gender, age, extrovert/introvert personalities also affect setting interpersonal distances [12]. Furthermore, it is well-known that different cultures hold different standards of personal and social space; for example, in Latin cultures these distances are usually smaller than in Nordic cultures.

However, despite the fact that the absolute distances depend on a number of factors, the space partitioning into concentric areas is common to all situations [38]; therefore, the categories of distances indicated by the study of proxemics can provide reference points for inferring the occurrence of social interactions and also benchmarks for assessing the accuracy of the system intended for interaction data collection.

3.1.2 Estimating distance between two mobile phones based on the analysis of radio signal strength

Existing solutions for distance estimation between two mobile phones exploit either acoustic components or mechanisms for transmitting/receiving radio signals. There has been only a few solutions based on the former approach which used ultrasound [68], which is not available in standard mobile phones, or acoustic signals emitted from the speaker [69], which require devices to be in earshot and in non-noisy environments. The current literature mostly reports the use of electromagnetic transmitting/receiving mechanisms to sense the presence of nearby mobile phones (such as Bluetooth scans [25][70]) or to infer the proximity based on co-location (such as NearMe [71]). However, both approaches have shown to provide the distance estimation accuracy in the order of 10 meters, which does not suffice for detecting the occurrence of social interaction on small spatio-temporal scale. With the respect to the study of proxemics, such a precision does not allow distinguishing distances linked to public and inner zones (related to social interactions).

The concept for estimating distance between two mobile phones, proposed in this thesis, is based on the RSSI analysis, which has been shown already to be a promising solution. RSSI based method is not limited to line of sight like infrared sensors are, and it is not privacy-sensitive in comparison to capturing audio data. In contrast to the approach of building a generic empirical model (regardless of the phone used, as implemented by Virtual Compass [10]), the approach proposed in this thesis maps RSSI values to distances relying on supervised learning, thus trading-off between the accuracy in distance estimation and the user effort in signal fingerprint collection. The reason for using a more costly method in terms of the end user effort is the fact that one of the pre-dominant factors affecting RSSI patterns is the receiver's characteristics [72] whose capturing can lead to a better system's accuracy. This hypothesis was tested in the experiments that follow, demonstrating that environmental factors have less prevailing impact on RSSI patterns than receiver's characteristics due to relatively short distances and no obstacles between receiver and transmitter. Unlike time-consuming measurements typically required for fingerprinting methods, the user effort is decreased to only a couple of minutes to calibrate the phone signal while achieving a comparable accuracy to full fingerprinting method.

The concept for estimating distance is tested using Wi-Fi signals. Nevertheless, other radio transmitting/receiving mechanism with accessible RSSI (such as FM or Bluetooth) available in mobile phones could be used for the same purpose or in combination with Wi-Fi.

3.1.3 Estimating distance between two mobile phones using Wi-Fi RSSI

Similar to indoor positioning systems that use fingerprinting technique, the method for distance estimation is based on analyzing RSSI values, observed on an unknown distance from the phone which transmits Wi-Fi signal (colloquially known as Portable Hot Spot or Personal Hot Spot). The receiving phone reports RSSI and estimates the unknown distance (Online phase - Figure 3.2) by applying the model built using a database that matches RSSI values with actual distances (Offline Phase – Figure 3.2). Considering the fact that RSSI patterns depend on a wide array of factors including receiver's characteristics and the type of environment, repeating the Offline phase would be required often to prevent accuracy degradation. Section 3.1.7 further discusses this issue and offers a solution for a fast calibration.

To acquire the training set, Wi-Fi signal was measured at different distances following a grid of 0.5 m while using two mobile phones (one in transmitting, the other one in receiving mode). Using smaller grid spacing for capturing RSSI pattern, which corresponds to acquiring more points in the training set, improves the system's performance but only up to a certain threshold when the accuracy starts to level off or even to degrade [72]. In the experiments, the grid of 0.5 m was found to be best suited considering the accuracy and the distances relevant for social interactions (Figure 3.1). For the same reasons, including measurements for distances greater than 8 m was superfluous.

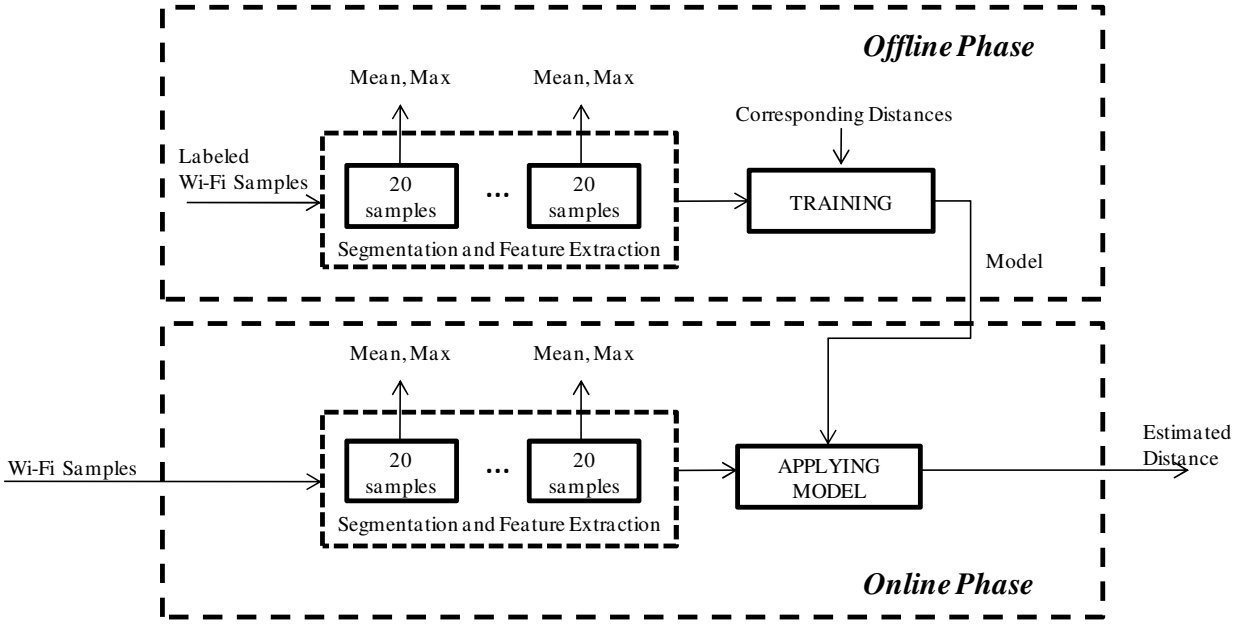


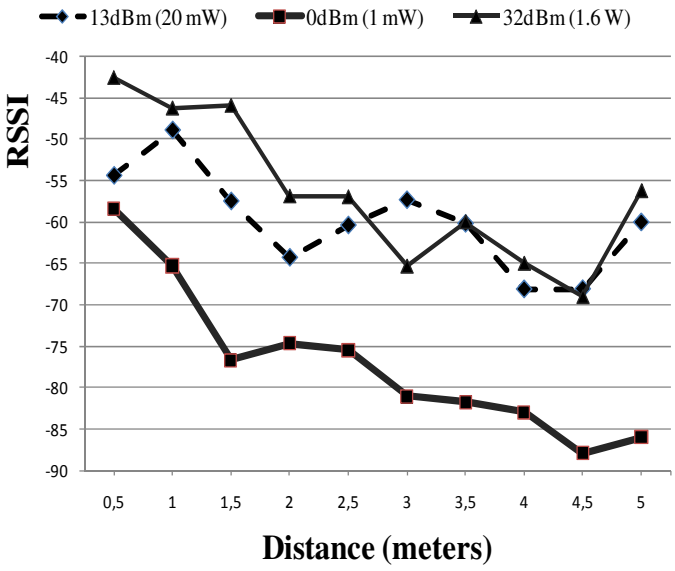
Figure 3.2: Block diagram of the interpersonal distance estimation

3.1.4 Feasibility of Wi-Fi RSSI pattern for distance estimation

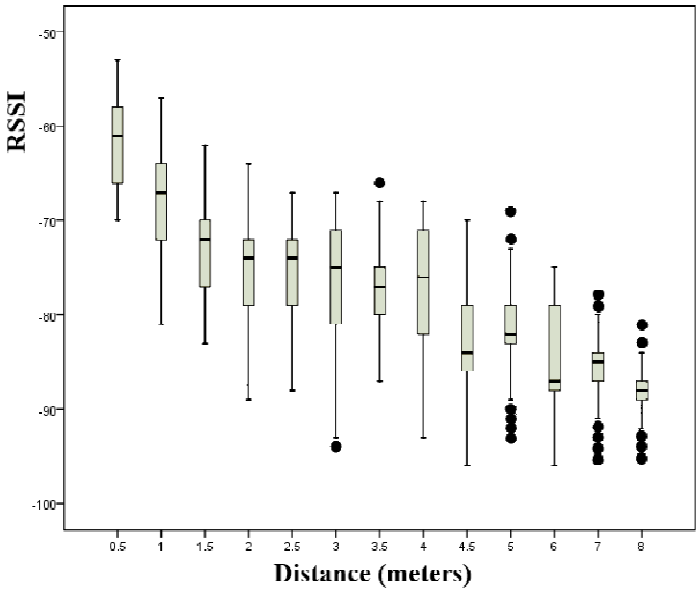
In order to evaluate the feasibility of distance estimation based on Wi-Fi RSSI, the RSSI dependence on distance was analyzed for three different transmitting power levels (Figure 3.3a): 32 dBm (1.6 W) – maximal available power level, 13 dBm (20 mW) and 0 dBm (1 mW) - minimal power level. The measurements were carried out in the same environment using HTC Desire mobile phone, while recording 300 samples with the sampling rate of 1 Hz for each of the distances following the grid of 0.5 m. The transmitting power of 0 dBm provided the smoothest and the most monotone characteristics (Figure 3.3a presents mean RSSI values) thus proving to be the best fit for short distance estimation. Afterwards, RSSI patterns were further analyzed setting the transmission power to 0 dBm, in seven different environments (indoor areas ranging from 30 m² to 90 m²) recording 300 samples every 0.5 m up to the distance of 8 m. The cumulative RSSI patterns are presented in

Figure 3.3b where small circles represent outliers, thick horizontal lines correspond to median values, bottom and top of each box corresponds to the first and the third quartiles of distribution and the whiskers extend up to 1.5 times the interquartile range (IQR). It can be seen that the RSSI shows relatively monotone characteristics across different environments while demonstrating the instability and fluctuations of the Wi-Fi signal, typically due to environmental factors [73]. Therefore, the distance estimation approach based on a simple RSSI threshold analysis (assigning ranges of

RSSI values to corresponding distances) did not suffice which led to pre-processing the signal and applying machine-learning techniques for distance estimation.



a) Three different power levels, one environment



b) Power level of 0dBm, seven environments

Figure 3.3: RSSI dependence on the distance

3.1.5 Estimating distance through classification and regression techniques

1. Data segmentation and features extraction

As the first step of data processing RSSI values were segmented by grouping every 10 consecutive samples (which, due to the sampling rate of 1Hz, corresponded to 10 seconds) and calculated signal characteristics for each group separately. The goal of grouping samples was to mitigate the effects of Wi-Fi signal instability in a short time frame [74]. Reducing the number of grouped samples resulted in the system's accuracy degradation while grouping more samples yielded no notable improvement in the distance recognition. Since RSSI distribution varies according to its mean [74], the mean value was selected as a candidate parameter to represent the RSSI pattern. It turned out that among other tested signal characteristics (such as standard deviation, minimum and median), the combination of the mean and maximal value provided the highest accuracy in distance estimation. Adding more parameters did not result in significant improvements of accuracy while, on the other hand, it increased computational requirements. Hence, every block of 10 consecutive RSSI samples (recorded over approximately 10 seconds) was represented in the training set with its mean and maximal value and was assigned to the corresponding distance.

2. Techniques Selection

Although the lognormal distribution of Wi-Fi RSSI was often assumed in the literature, it only represents a part of real RSSI distributions [74]. Therefore, Naïve Bayes with Kernel Density Estimation (KDE) classification was suitable choice considering the fact that it stands for a flexible nonparametric technique; it was also evidenced in the experiments to provide the best accuracy. However, several classification techniques that were tested demonstrated similar performance in distance estimation; as an illustration, the results for Linear classification are reported as well.

On the regression side, Gaussian Process (GP) is a well-suited regression method for localization for the following reasons: a) it does not require a discrete representation of an environment, b) being non-parametric approach it provides adequate approximation for a wide range of non-linear functions, c) GP parameters can be estimated from training data applying well-known algorithms [75].

3.1.6 Experimental setup and results

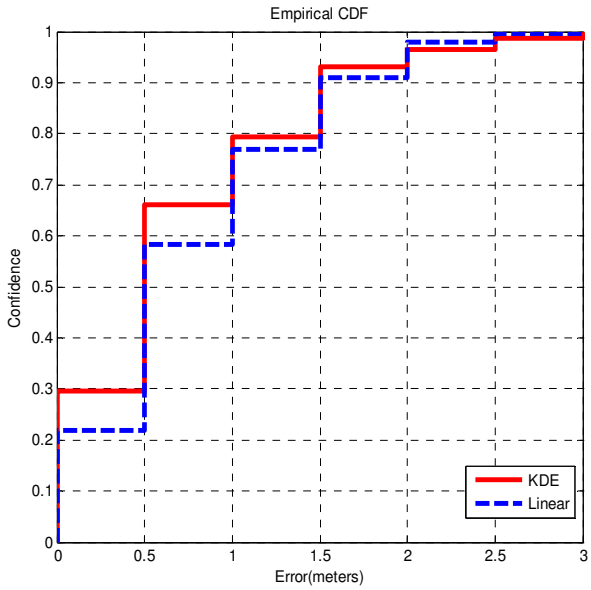
The testbed consisted of six mobile phones (with Android operating system) including three different models, namely HTC Desire, Samsung Nexus S, and HTC Nexus One that were modified to allow adjustment of transmitting power. Distance estimation accuracy results were consistent for all tested phone models across six environments; evaluating performance for other phone models is out of the scope of this work; however, based on the phone models already tested, large disparities are not to be expected.

Measurements were taken in three offices with dimensions of 12x8 m, 6x5 m and 6x3 m, a balcony of 12x2.5 m, a meeting room of 10x8 m and outdoors in front of the building. For testing the system's accuracy, a pair of phones was used – one in transmitting and the other one in receiving mode. Following a grid of 0.5 m, RSSI was measured for 5 minutes on each distance between phones starting from 0.5 m to the point in which either signal degraded to its minimal level or it was the furthest accessible point within room dimensions. The maximal distance in the experiments was between 5 and 8 m thus covering all the distances relevant to the study of proxemics [76][19]. The measurements were stored locally on the phone memory but for facilitating data analysis uploaded on the server during the evaluation period. The standalone implementation is presented in [77].

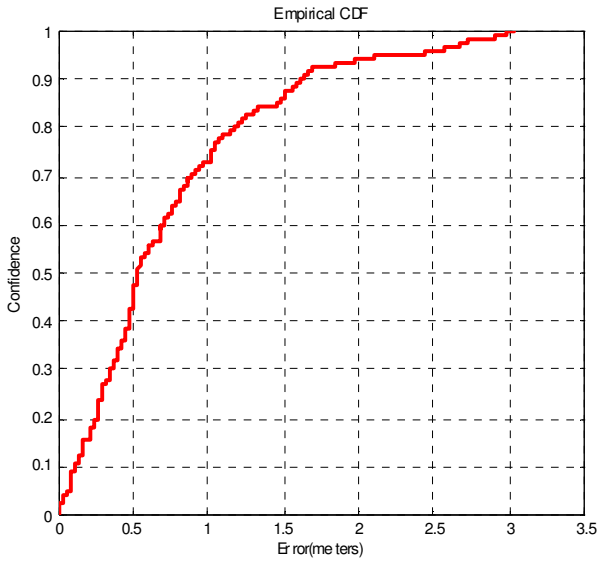
The accuracy was estimated in the offline phase by applying a cross-validation method: RSSI pattern captured in one out of six environments (outdoor, three offices, balcony and a meeting room) was used for building the model (i.e. as a training set) while measurements from five remaining environments were used for testing. In this manner, the procedure was repeated to cover all the combinations regarding distinct training and test sets across six environments. The RSSI characteristics were calculated over every block of 10 samples and queried separately to estimate an unknown distance. The cumulative distribution function of the distance estimation error was plotted to evaluate the system's accuracy.

Figure 3.4 shows the system's accuracy in the case of using the same phone (same model) acting as a receiver in both training and test phase. The median estimation error (50th percentile) of approximately 0.5 m was achieved using all the three applied machine learning techniques. Naive Bayes with KDE showed a slightly better

overall performance, providing distance estimation with 50th percentile error of 0.5 m and 90th percentile error of 1.5 m. It should be mentioned that the system's accuracy did not significantly differ when tested separately in outdoor conditions; therefore the results for this case were not reported independently.



a) Classification



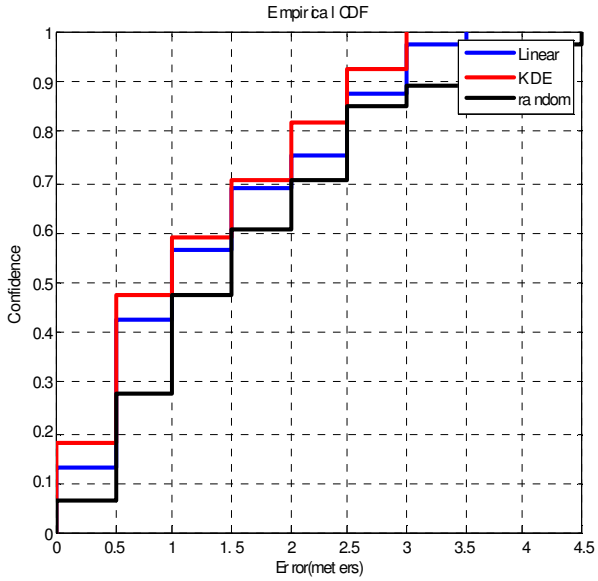
b) GP Regression

Figure 3.4: Cumulative distribution function of the distance estimation errors (same receiving phone for training and testing)

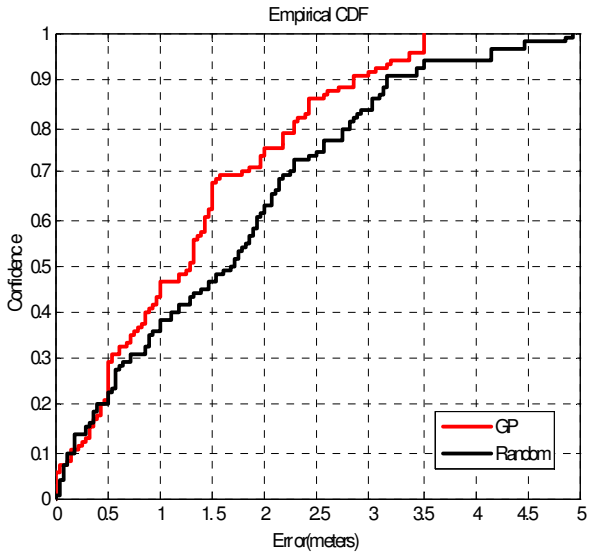
When different models of phones were used for training and test phase, the system's accuracy significantly degraded (Figure 3.5). This is due to the fact that

RSSI patterns highly depend on the receiver characteristics [72] which are likely to be different across different phone models. Figure 3.5 presents distance estimation accuracy plotted against the baseline performance represented as a random estimation out of the set containing distances from 0 to 5 meters. The median error was approximately 1 m while 90th percentile error was between 2.5 m and 3 m. However, considering the main goal of recognizing distances related to social interactions, the system did not provide satisfactory accuracy in this case. The next section further discusses this issue.

The change of a mobile device, which acts as a hot-spot while keeping the same transmitting power did not result in significantly different RSSI patterns; therefore the results for this case were not reported separately.



a) Classification



b) Regression

Figure 3.5: Cumulative distribution function of the distance estimation errors (different receiving phone for training and testing)

3.1.7 Fast calibration based on propagation model

Having a generic application for inferring social interactions would require a generic radio propagation model that would reflect RSSI patterns considering various models of mobile phones and environmental conditions. However, the possibility of creating such model, which would provide a reasonable system's accuracy for a number of phone models, is limited given that the range of RSSI values is a function of effective transmitter power, path loss, receiver antenna gain, and receiver sensitivity [72]. In addition, fluctuations of RSSI may be caused by environmental factors including furniture layout in the room of interest, air temperature, humidity and presence of people [73]. Since some sources of uncertainty are difficult or impossible to be taken into account, the main goal becomes identifying pre-dominant factors that influence RSSI patterns and modeling their effects.

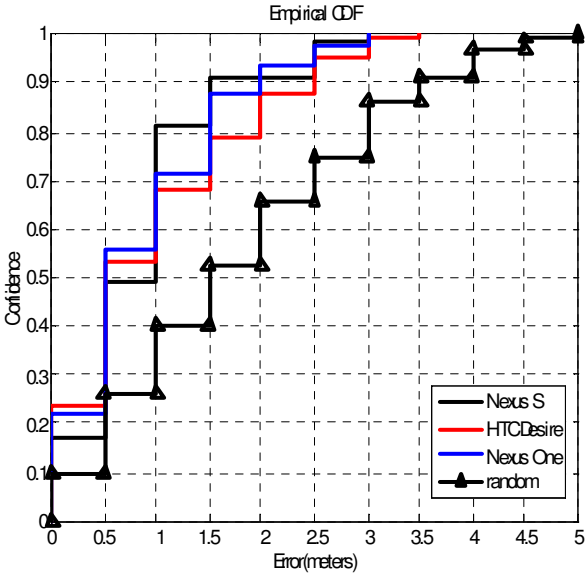
Considering the fact that other conditions in the experiments remained constant, the main cause of degradation in the case of using different phones for model building and for distance estimation (the difference in results shown in Figure 3.4 and in Figure 3.5) lies in the change of receiver's characteristics. The application should keep the transmitting power constant to 1 mW, however the remaining issues to be addressed are related to receiver characteristics and path loss. This would require a calibration i.e. acquisition of a RSSI pattern database for any new model of the phone

intended to use the application. Repeating RSSI measurements would be laborious and time-consuming process since achieving satisfactory accuracy requires a number of distances (in this case following a grid of 0.5 m). To avoid this, a proposed solution is based on calibrating only one point by measuring RSSI for a couple of minutes on a fixed distance of, for instance, 1 m. Once the RSSI is captured, the rest of the training set is estimated applying the following propagation model [78]:

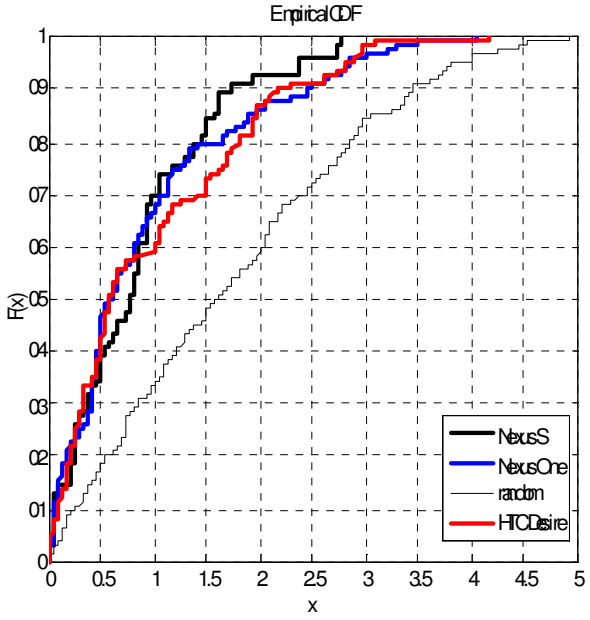
$$P(d)[dBm] = P(d_o)[dBm] - 10n\log(\frac{d}{d_0}) - X \qquad (3.1)$$

where n is the path loss exponent, P(d0) is the signal power at the reference distance d0 from the transmitter phone (in this case 1 m) and d is the distance in which RSSI is estimated by applying the model. X is a component that reflects the sum of losses induced by each wall between the transmitter and receiver. It was found empirically from the training sets that the best suited value for the coefficient n is 1.5, while for X is zero (there are no walls or other obstacles between points).

This method was evaluated by measuring RSSI on the distance of 1m in one out of six environments, generating the rest of training set by applying the propagation model and assess the accuracy using RSSI measurements from five remaining environments. The procedure was repeated for all the environments and overall performance is plotted in Figure 3.6 as a cumulative distribution function of distance estimation errors (for each of the phone model that was used).



a) Classification



b) GP regression

Figure 3.6: Distance estimation accuracy using training set generated by applying the propagation model

This demonstrates that similar performance as in the case of acquiring a full training set can be achieved by investing a minimal effort of performing the calibration for a couple of minutes (Figure 3.4 and Figure 3.6). All tested models showed similar performance - the median accuracy of approximately 0.5 m and 90th percentile error between 1.5 m and 2.5 m.

Unexpectedly, calibrating the phone and testing in the same environment provided similar accuracy as in the case of performing calibration and testing in different environments (which was evidenced across all six environments). This may be indicative that the pre-dominant factor that influence RSSI pattern lies in receiver's characteristics. Less prevailing impact of environmental conditions may be explained by relatively short distances and no obstacles between receiver and transmitter which could affect the signal propagation. This was further evidenced in the experiments of real-life settings conducted in a wide array of environments and different phone models which is presented in chapter 5 and chapter 6.

3.1.8 Accuracy in distinguishing classes of distances related to proxemics

Previous sections (3.1.6 and 3.1.7) provided the evaluation of the system's absolute accuracy. By mapping of estimated distances into different categories of social interactions through the study of proxemics, this section assesses the accuracy in distinguishing interpersonal distances with respect to personal, social and public space. The intimate space which includes distances up to 0.45 m cannot be reliably detected using the proposed method, thus all the recognized distances below 1.2 m were categorized as a personal space. The reason lies in the fact that the detection of such short distances between people is highly affected by the place of carrying the phone (such as a pocket, a case or a bag). The distance estimation accuracy is broken down into the three classes – personal, social and public space and the results are presented in Table 3.1 in the form of a confusion matrix.

Table 3.1: Break-down classification accuracy related to the categories of interpersonal distances defined by the study of proxemics

Ground-truth	a) Same phone for training/test			b) Calibration method		
	Personal	Social	Public	Personal	Social	Public
Personal	81%	19%	0%	81%	19%	0%
Social	0%	67%	33%	28%	51%	21%
Public	0%	17%	83%	2%	14%	84%

Distances related to personal and public space were recognized in more than 80% of cases both when a) performing training and test procedure with the same phone and b) using calibration method (Table 3.1). Distinguishing social space from personal and public space resulted in lower accuracy, 67% and 51% with respect to different methods for acquiring the training set.

According to the study of proxemics, distances related to personal and social space are used by subjects for different types of social interactions thus distinguishing these two categories from public space would correspond to inferring the distances relevant for social interactions in general. Table 3.2 presents the results when the system's accuracy is broken down further in two groups – social interactions related distances and public space distances. In 82% - 86% of cases the system successfully distinguished the two groups.

Table 3.2: Recognizing two groups of distances related to social interactions and public space

Ground-truth	a) Same phone for training/test		b) Calibration method	
	Social	Public	Social	Public
Social Interaction Distances	82%	18%	86%	14%
Public Space	17%	83%	16%	84%

Spatial settings in social interactions can be affected by a number of factors thus the boundaries of personal, social and public space may vary across different testbeds. However, once absolute interpersonal distances are estimated, it becomes trivial to adjust their categorization according to the culture where the experiments were conducted or some other factor that affects setting interpersonal distances.

3.1.9 Comparison of distance estimation systems

Table 3.3 shows related systems for peer-based distance estimation that are reported in the current literature.

Table 3.3: Comparison of proximity/distance estimation systems

Project	**Accuracy**	**Method**
Virtual Compass [65]	50^{th} percentile error: 0.9m, 90^{th} percentile: 2.7m	Wi-Fi + Bluetooth
BeepBeep [69]	50^{th} percentile error: within 2cm	Acoustic-based
NearMe [71]	RMS error: 10m-20m	Comparing Wi-Fi fingerprints
Relate System [68]	50^{th} percentile error: 2cm – 4cm	Ultrasound

Relate System [68] calculates the relative position of devices relying on a custom ultrasound hardware. This approach provides a very accurate estimate of distance, with the median accuracy in the order of centimeters, but it requires ultrasound emitters/receivers that are not available in standard mobile phones. Moreover, techniques that rely on ultrasound or detection of the phase offset of transmitted radio waves are difficult to implement using the hardware and APIs available on commodity mobile phones [65].

NearMe compares clients' list of Wi-Fi access points and signal strengths to compute the distance between devices. Unlike localization system based on Wi-Fi fingerprints, NearMe does not rely on calculating an absolute location thus it requires no calibration and minimal setup. This method achieves relatively low accuracy in comparison to other systems with an RMS (Root Mean Square) error of 10 to 20 meters.

BeepBeep [69] is a highly accurate acoustic-based system for estimating distance between devices which requires only a set of commodity hardware – a speaker, a microphone and a form of device-to-device communication. Each device emits a sound signal and collects its own signal and a signal from its peer. Distance estimation is based on counting the number of samples between these signals and exchanging the time duration with its peer thus calculating two-way time of flight. The approach requires wireless communication to coordinate devices and to exchange the time duration. Noisy environments impact the accuracy of the system while the devices that are not in earshot cannot be detected; this limits applicability for mobile phones considering the fact that they are typically carried in places that affect sound propagation including pockets, cases and bags.

Similarly to the distance estimation system proposed in this thesis, Virtual Compass [65] exploits transmitting mechanisms embedded in mobile phone and RSSI analysis. Mapping RSSI to distance was performed with empirical propagation models enhanced by incorporating the uncertainty which provided the average accuracy of 3.4m and 3.91m when Bluetooth and Wi-Fi (respectively) were tested separately using nine devices in a 100m² indoor area. The fusion of the two transmitting mechanisms achieved the median error of 1.41m for nine devices while in the case of two devices in the same area the median error was 0.9m and the 90th percentile error was 2.7m. In comparison to the distance estimation method proposed in this chapter, the advantages of Virtual Compass includes estimating positions of devices in 2D plane, algorithms for energy efficient use and not requiring training phase. However, the system proposed by this thesis provides a higher accuracy with the use of solely Wi-Fi, does not require communication between devices and broadcasting the distance to each of peers, while training phase is facilitated with a fast calibration method which makes the approach adaptive to different applications, environments and phone models.

3.2. Estimating relative body orientations

3.2.1 Relative body orientations and social interactions

In addition to the interpersonal distance, posture is also often investigated measure of nonverbal behavior in social interactions, typically indicating a communi-

cator's attitude toward his peer [79]. One important dimension of posture, the relative body orientation, represents the angle between the torso's orientations. It can reflect the level of immediacy in face-to-face interaction; in addition, several experiments demonstrated that, in the case of female individuals, a more direct position indicates a more positive attitude [79]. Nevertheless, a direct face-to-face position refers to the need of continuous mutual monitoring and this type of interaction is usually more active [9]. When two subjects hold more parallel position, this may be a sign that they are either buddies or less mutually interested [9]. In addition to interpreting subjects' attitudes, the relative body orientation shows the high predictive power for recognizing the occurrence of an ongoing social interaction, when combined with the interpersonal distance [12].

3.2.2 Estimating relative body orientation

Relative body orientation refers to the angle between the orientations of torsos [12] considering two subjects that are facing each other. The solution proposed in this thesis makes the use of the orientation sensor embedded in mobile phones in order to estimate the relative body orientation of subjects. The orientation sensor reports the following values (expressed in degrees):

Azimuth – the angle between the magnetic north direction and the y-axis, around the z-axis (0° to 359°); 0° corresponds to North, 90° to East, 180° to South, and 270° to West.

Pitch – the rotation around x-axis (-180° to 180°) with positive values when the z-axis moves towards the y-axis.

Roll – the rotation around y-axis (-90° to 90°) with positive values when the x-axis moves towards the z-axis.

Knowing the relative position between the body and the phone orientation is the fundamental condition in order to recognize the individual's body orientation and the relative body orientation between subjects. Once this relationship is determined, calculating the relative body orientation would require relative processing of azimuth, pitch and roll values. In the experiments that follow, participants were instructed to carry the phone with a display facing outside in a case, placed on the right side of the hip. Depending on the body posture the phone could mostly rotate around y- and z-

axes thus considering the sum of azimuth and roll was sufficient to compare the body orientations of subjects (being in a case, the display plane was always parallel to the body implying little or no rotations around x-axis). Evidently, the position for carrying phone is by no means limited to the hip but the constraint relates to the constant position with respect to the body and must be known to the algorithm, in order to calculate relative body orientation. However, in the recent study, Shi et al. [80] demonstrated that it is possible to automatically detect on-body position of the mobile phone by utilizing the fusion of accelerometer and gyroscope. Furthermore, chapter 5 and chapter 6 demonstrate that without knowing the position of the phone, meaningful and informative social interaction features can be extracted from orientation sensor data both for inferring the occurrence of social interaction and for social context analysis.

It would be superfluous to test the accuracy of the sensor itself for each phone model that was used in the experiments as it is expected to be done at phone/sensor producer. Instead, the relationship between the estimated relative body orientation and the ground-truth position was analyzed during face-to-face social interaction. In this case, the ground-truth refers to the position perceived by subjects, for instance as a direct face-to-face or parallel. Considering the approach of carrying the phone in a constant place on the body, the uncertainty in estimating the subject's orientation is introduced by a human factor, which in this case particularly influences positions that subjects hold while interacting and placing the phone exactly on a pre-defined spot. Two sets of controlled experiments were conducted: one consisted of scheduled meetings involving overall six different subjects (two of them meeting at time, both standing) and the other one was conducted in two sessions where each session included four people that were asked to talk for 20 minutes in the meeting room, constituting one-on-one interactions. There were overall 31 and 6 recorded social interactions respectively with the duration of 3.6±3.4 minutes (varying from 1 to 10 minutes). The subjects, not introduced with this study, were asked to interact in a face to face formation and to label each single social interaction by annotating using the start and the end button in the application.

The phones were sampling orientation values with the frequency of 1Hz and were uploading to the server. As in the case of sampling Wi-Fi for estimating interpersonal distances, uploading data to the server was done for facilitated data analysis (the

stand-alone application which allows exchanging the information of absolute orientations without establishing a connection is described in [77]). The individual's orientation (that corresponded to the sum of azimuth and roll in this case) was averaged every frame of 10 seconds i.e. 10 samples while the relative body orientation of subjects was calculated only if the standard variation of the samples was less or equal to 10 degrees for each subject, otherwise the current frame of samples was taken out of the consideration. This was done for indicating the moments in which subjects do not hold a stable orientation due to body movement thus eliminating this source of uncertainty for relative body orientation estimation. The threshold of 10 degrees was selected through a trade-off between decreasing the threshold of standard deviation of the estimated relative body orientation and decreasing the amount of discarded data (proportional to increasing threshold).

Figure 3.7 shows the distribution of the estimated relative body orientation during 37 face-to-face social interactions observed in controlled experiments, where 180 degrees between hip (i.e. torso) lines corresponds to a direct face-to-face formation.

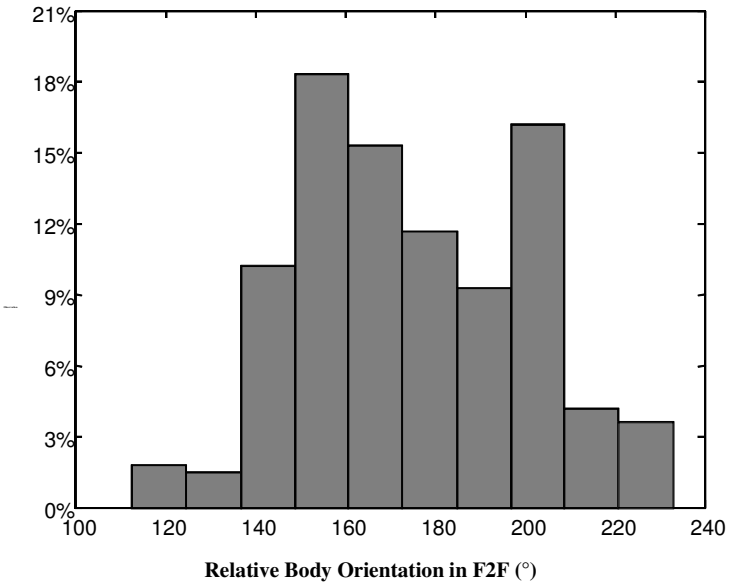


Figure 3.7 Relative body orientation in two-person meetings (degrees)

The mean value of angle was 178° with the standard deviation of 25°. Overall 20% of data was discarded due to not stable orientation according to the threshold of 10° described above. Such result indicates that estimated relative body orientation reflected the ground-truth position, which in this case was the relative body orientation

in direct face-to-face interactions. It was not expected that during two-person face-to-face interactions the relative body orientation will be constantly 180°, thus the standard deviation of 25° may be taken as the acceptable one, mirroring the ground-truth position.

3.3. Summary

The way how people use and organize space they share with others in social interactions, reflects their social and emotional closeness, attitude, cultural background, type of personality and other similar phenomena. At a more basic level, setting appropriate spatial settings is a necessary condition for accomplishing a co-located face-to-face interaction. Therefore, inferring spatial settings represents an important aspect when it comes to monitoring of social interactions.

This chapter investigated the possibility of using the mobile phone as a source for collecting information about interpersonal distances and relative body orientations, parameters which describe spatial settings between subjects. Using the RSSI analysis of the radio signals transmitted from one and received by the other phone, the proposed method estimates the distances between users, assuming that they carry mobile phones, with the median accuracy of 0.5m (demonstrated with Wi-Fi signals). This allowed distinguishing distances linked to public and social zones according to the study of proxemics with the accuracy of above 80%. In order to estimate relative body orientation, the method relies on the compass sensor embedded in modern mobile phones assuming the knowledge of the position where users carry their mobile phones. The on-body position of the mobile phone can be either reported by subjects or automatically detected through the existing algorithms. However, chapter 5 and chapter 6 demonstrate that without knowing the position of the phone, compass sensor data can be used for detection of face-to-face interactions and for the social context analysis.

The fact that people habitually carry the mobile phone makes this device a suitable source for unobtrusive and continuous monitoring of social interactions..

Chapter 4

4. Speech Activity Detection

Although people, consciously and unconsciously, communicate in a nonverbal way, speech is still considered to be the main modality of a conversation and its direct manifestation [22]. Looking from the perspective of a human observer whose task is to collect interaction data, annotating the occurrence of a conversation pertains to witnessing the speech activity, while most of sensor-based systems for detecting social interactions rely on the audio data analysis. In certain situations nonverbal cues, such as interpersonal distance and relative body orientation, can suffice to recognize whether or not a social interaction takes place (which is further analyzed in chapter 5), however in order to ascertain the occurrence of a conversation, it requires also the knowledge about speech activity [8]. For example, two colleagues can have desks one in front of another thus being physically close and facing each other the whole working day while not communicating; on the other hand, colleagues may have a conversation at the office while facing monitors. Interactions detected from speech and collocation have been shown to be different [13] which highlights the need of having knowledge about speech activity in order to reliably detect an ongoing conversation.

In the technological community, speech activity has been pre-dominantly recognized using microphones, either mounted in the area of interest or embedded in a mobile device (the mobile phone [13] or specialized device such as Sociometer [8]). As discussed in chapter 2, the limitations of these approaches include 1) sensitivity to false positives – nearby conversations can be unintentionally picked up;, 2) privacy and ethical issues – in a number of situations (for example, in public spaces or in the case of monitoring patients) audio data cannot be obtained due to legal or ethical norms [35]. Even when applying privacy sensitive recording techniques, activating microphone may cause changes in natural behavior of subjects if they perceive it as privacy intrusive. Therefore, a few alternative methods that aimed to infer speech activity are based on mouth movement, fidgeting, or gestures [35] [81] detected using

machine vision. However, this limits application scenarios to limited areas that are covered with the camera system. These reasons prompted the investigation presented in this chapter with the goal of providing an alternative to microphone-based speech detection method commonly used by the systems for sensing social interactions, while still allowing a continuous data collection (not limiting the application only to certain areas).

The proposed method for speech activity detection is based on identifying another manifestation of speech different than voice, namely the vibration of vocal chords. The phonation-caused vibrations spread from the area of larynx to the chest level, representing the exhibition of speech activity which can be automatically detected through the use of accelerometer. In the domain of speech analysis, non-acoustic sensors have been used so far to investigate speech attributes [82] [83], speech encoding [83], and to augment communication possibilities in patients with special needs [84], however no work has relied on accelerometers to detect the status of speech in social interactions. The accelerometer-based approach does not require sensitive data thus it is not expected to face ethical issues or privacy concerns in comparison to microphone based approaches.

4.1. Methodology

Vocal chords (also known as vocal folds) are muscles within larynx that vibrate when air from lungs passes through thus producing voice [85]. The fundamental frequency of vocal chords vibrations depends on a variety of factors including age, gender and individual differences [86]. After the age of 20 the predicted fundamental frequency of vocal chords remains approximately 100Hz for male and 200Hz for female adults [86]. Therefore, identifying vibrations of these fundamental frequencies produced by vocal chords during phonation pertains to speech activity detection. Instead of a purpose-built accelerometer (with an appropriate shape, targeted frequency range and sensitivity), this chapter investigates the use of an off-the-shelf accelerometer thus aiming to an easily applicable and a cost effective solution. Since mounting sensors on the neck (close to the larynx area) may be obtrusive, the chest surface was selected as a suitable body position, which is already being used for placing various

sensors including cardio, respiratory and kinematic sensors. Sundberg [82] identified a number of factors that contribute to the chest vibrations during phonation and examined the distribution of displacement amplitude over the chest wall surface, demonstrating that the vibrations can be detected all over the chest with the highest displacement amplitude located in the central part of the sternum, which is the area chosen to place the sensor on (Figure 4.1).

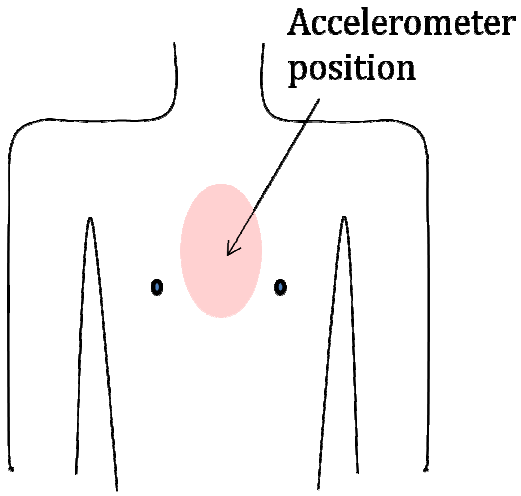


Figure 4.1: Area on the chest for placing accelerometer

4.2. Accelerometer-based approach to recognize speech

The concept of using an accelerometer for recognizing speech activity is based on detecting phonation-caused vibrations at the chest level, targeting frequency range approximately between 100Hz and 200Hz. On the other hand, it is important to examine if there are potential sources in everyday life that produce components in the same range of frequencies which can be confused with speech activity. It can be noted that daily physical activities are not expected to overlap with vocal chords vibrations in the frequency domain since they typically occupy frequency ranges lower than 20 Hz [87]. However, this investigation is focused on the following two aspects:

1) Whether the characteristics of off-the-shelf accelerometers (i.e. not specifically designed for detecting small vibrations) are sufficient for recognizing speech activity and discriminate it from other components in the frequency spectrum. This concern refers mostly to low amplitudes of the chest wall vibrations [82] that may be

similar to noise level, imperfect contact between the sensor and the chest, and physiological and acoustic differences between genders [86] and also across all individuals.

2) Whether other sources of vibrations encountered in everyday life including elevator, car, bus, train or airplane, whose engines provide components in higher frequency ranges that may result in false positives for speech detection.

To evaluate the approach of detecting speech activity based on analyzing frequency spectrum of data acquired from an off-the-shelf accelerometer [88] attached to the chest (Figure 4.1). The specifications of the accelerometer (not specifically adapted to detect small vibrations) are the following: the range of ±1.5 and ±6g, sensitivity of 800mV/g at 1.5g and a maximal sampling rate of 512Hz. According to the Nyquist-Shannon sampling theorem, the ceiling boundary frequency component that can be detected using this accelerometer is 256 Hz, which fulfils the requirements for the intended application (since the fundamental frequencies of vocal chords are approximately 100Hz for males and 200Hz for females). To analyze the frequency domain of acceleration time series (square roots of the sum of the values of each axis *x, y* and *z* squared), the method relied on Discrete Fourier Transform (DFT) defined for a given sequence x_k, k = 0, 1, … N-1 as the sequence X_r, r = 0, 1, … , N-1 [89]:

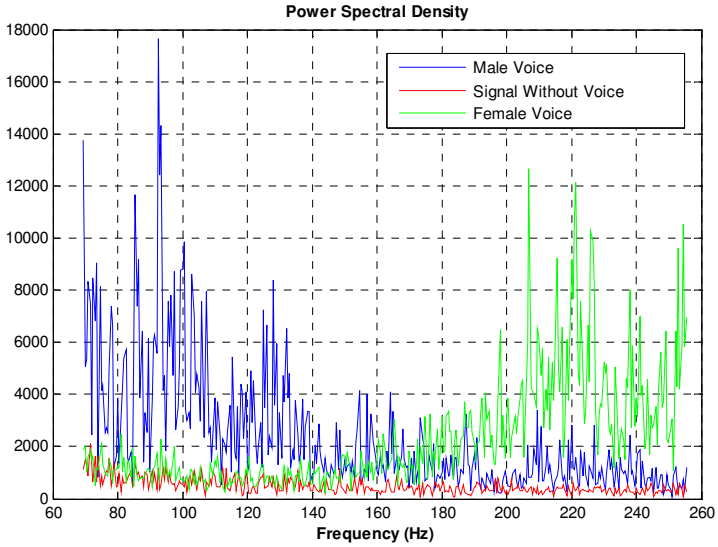
$$X_r = \sum_{k=0}^{N-1} x_k e^{-j2\pi rk/N} \quad (4.1)$$

Frequency spectrum was analyzed in MatLab applying the Fast Fourier Transform (FFT) to calculate the DTF and then the power spectral density was computed.

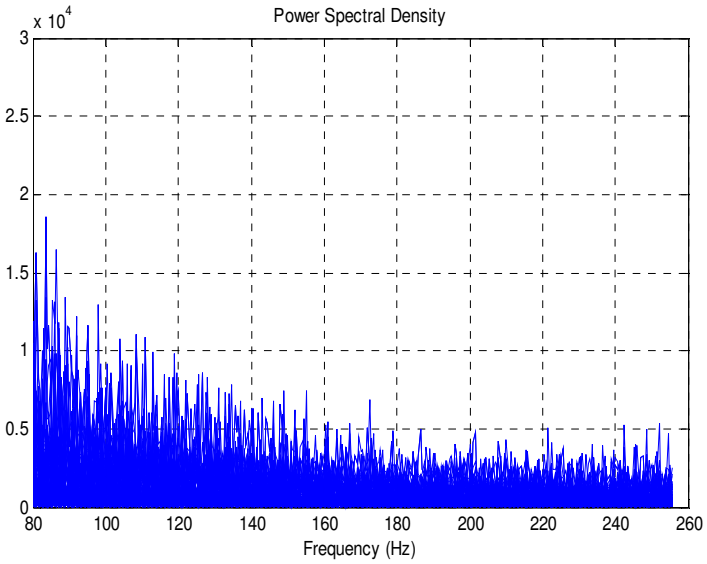
As expected, low amplitudes of the chest wall vibration were similar to the noise level thus only by analyzing the frequency spectra was not possible to distinguish accelerometer readings that contained speech from those that contained noise. In order to tackle the problem of noise, a noise cancelling strategy [90] was applied which consists of summing frequency spectra in time. This strategy is based on the assumption that the signal components are always focused in the same frequency range in contrast to noise that is, in this case, more random. Considering time frames for performing power spectral density analysis, the best accuracy was achieved by analyzing a sum of power spectral densities computed separately for five consecutive 2-second long time series (corresponding to 1024 samples in this case). Hence, each 10-seconds

frame was represented with the power spectral density that was a sum of spectral densities computed for each 2 seconds. Therefore, the goal was to recognize the presence of spectral components that correspond to speech with the resolution of 10 seconds. Processing data in 10-second time frames resulted in the highest accuracy regardless of the duration of the speech i.e. whether there was only one word spoken or a continuous talk of 10 seconds. Decreasing the resolution corresponded to lower ratio between speech amplitudes and noise levels while processing data in longer time units was more likely to fail in detecting shorter durations of speech.

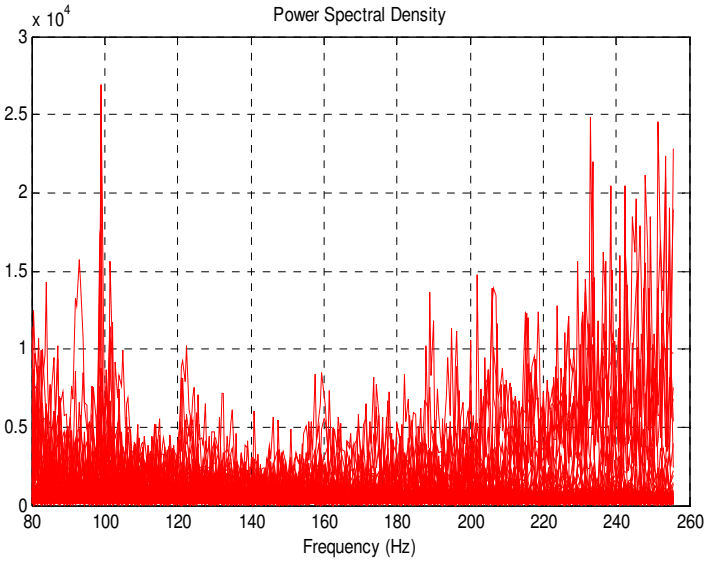
Figure 4.2a shows distinct examples of frequency spectra for 10-second samples containing no voice, male voice and female voice where readings without voice correspond to a stable position of accelerometer (i.e. of a subject wearing accelerometer) without speech activity. However, even mild activities (such as moving in the chair, small movements while standing or walking) resulted in noise in higher frequency ranges (Figure 4.2b) making the frequency spectra similar to the one related to speech (Figure 4.2c). Figure 4.2b presents a set of frequency spectra that contains overall 3 hours (separated in 10-second time frames) of different activities including sitting, standing and walking without speech. A set of frequency spectra corresponding to speech activity of 11 male and 10 female subjects is shown in Figure 4.2c (10-second time frame of continuous speech for each subject).



a) Single readings - Male/Female/No Voice



b) Set of readings that do not contain speech activity



c) Set of readings that contain speech activity

Figure 4.2: Power spectral densities of accelerometer readings

Visualization of frequency spectra evidenced the non linear nature of the classification problem; therefore, several classification techniques (namely SVM, Naïve Bayes, Naïve Bayes with kernel density estimation and k-NN) were considered being evaluated using various features for characterizing the spectral density (namely mean, maximal, minimal, and integral values regarding different frequency ranges). It turned out that Naïve Bayes with kernel estimator applied on the two parameters – integral and mean values of the components between 80 Hz and 256 Hz, provided the highest classification accuracy.

Here no further elaboration is provided on the classification selection, a choice of signal parameters, frame size for calculating power spectral density and the resolution since they strongly depend on the accelerometer's characteristics. However, considering the predicted fundamental frequencies of vocal chords, the requirement for the accelerometer is sampling rate of minimum 200 Hz for detecting male voices and 400 Hz for detecting female voices. The accuracy of the approach is provided in the following,.

4.3. Experiments and results

In total, 21 subjects participated in the speech activity detection experiment (11 males, 10 females; 31.8±7.6 years old). Each subject was asked to read out loud the article from the latest newspapers for at least two minutes, while having the Shimmer accelerometer attached to the chest with an elastic band (Figure 5.5). The approach was evaluated separately for each subject through cross-validation of two sets, one including the frequency spectra of 10-second frames containing subject's voice and the other including only spectra of accelerometer data samples recorded during mild physical activities without voice. 10 out of 11 male and 9 out of 10 female voices were successfully recognized, demonstrating that in large majority of the cases the accelerometer was sufficient to distinguish the frequency spectra of readings with and without voice despite the imperfect skin-sensor contact and individual subjects' characteristics.

In addition, a set of accelerometer data was created containing speech activity of 19 subjects excluding 2 subjects that were not previously detected (overall, 2 minutes each subject, that is 38 minutes, divided in 10-second time frames) and accelerometer readings that contained physical movements without voice (approx. 2 hours of accelerometer readings that included sitting, standing and normal speed walking in 10-second data resolution). This was done so that a generic speech detection model can be built. The voice recognition accuracy was estimated through leave-one-out method of sequentially selecting accelerometer readings that corresponded to one subject/one activity as a test unit while using the rest of the set for building the model (training set for Naïve Bayes with KDE classification). The voice was correctly rec-

ognized in 93% of cases while mild physical activities without voice induced false positives in 19% (Table 4.1a). The same model was used to test accelerometer readings acquired in more intensive activities such as fast walking or running which resulted in 29% rate of false positives (Table 4.1b). Furthermore, the goal was to investigate whether some sources that may be encountered in everyday life including elevator (5min of data), car (30min of data), bus (30 min of data), train (20min of data) or airplane (1 hour of data) whose engines provide components in higher frequency ranges result in false positives in speech detection. It turned out that elevator, train and airplane do not present an additional issue for the speech recognition, causing the same rate of false positives as physical movements performed in normal conditions (Table 4.1a and Table 4.1c) while travelling in a car or a bus increases the occurrence of false positives to the rate of 32%. Intense physical activities and the transportation vehicles did not affect the recognition of speech i.e. the rate of true positives and false negatives remained above 90%.

Table 4.1: a) Voice/mild activities; b) Voice/intense activities, c) Voice/source of higher frequencies

a)	Voice Detected	No Voice Detected
Voice	93%	7%
Mild Activities	19%	81%

b)	Fast Walking or Running
No voice detected (true negatives)	71%
Voice detected (false positives)	29%

c)	Elevator	Bus/ Car	Train	Airplane
No voice detected (true negatives)	80%	68%	81%	79%
Voice detected (false positives)	20%	32%	19%	21%

The results demonstrate that the speech activity can be reliably detected in typical daily situations that include mild activities. More intense activities, such as running, and certain types of vehicles, such as car or bus, may result in a higher rate of false positives. Nevertheless, this may be mitigated by using a different type of the accelerometer.

In addition to phonation there are other causes of vocal chords vibrations, which can be incorrectly classified as speech activity such as coughing or mumbling. Certainly, this should be considered when designing a system intended for users that prevalently suffer from such symptoms. However, in healthy subjects their occurrence is less frequent and typically negligible in comparison to speech.

4.4. Summary

The evaluation presented in this chapter demonstrated the possibility of using an off-the-shelf accelerometer for inferring speech activity. This method is based on identifying vibrations caused by vocal chords during phonation which belong to the frequency range approximately between 100Hz and 200Hz. Instead of a purpose-manufactured accelerometer (with an appropriate shape, targeted frequency range and sensitivity), this chapter investigated the use of an off-the-shelf accelerometer which affords an easily applicable and a cost effective solution. The results demonstrated that the accelerometer attached at the chest level can distinguish speech activity from mild daily activities with an accuracy of 90%, while more intensive activities and certain vehicles may result in higher rate of false positives, up to 32%. However, these issues can be mitigated by using an off-the-shelf accelerometer with different characteristics.

The accelerometer-based approach provides an alternative to the speech detection based on audio data analysis which is prone to provoking privacy concerns in subjects. Besides, capturing audio data may be considered unethical and illegal in a number of situations. Using an accelerometer to identify physiological effects of phonation does not require capturing sensitive information thus overcoming the above-mentioned drawbacks of microphone-based methods. Such an approach faces the challenges of interpreting noisy data acquired from a source which is not dedicated for speech activity detection. Moreover, wearing a sensor at the chest level may be perceived as obtrusive and consequently it may stigmatize monitored subjects. However, wearing comfort issue is expected to be mitigated considering the wide variety of sizes and shapes of the currently available accelerometers,

Chapter 5

5. Detecting Social Interactions

Through several studies, Groh et al. developed probabilistic models for detecting social interactions based on various nonverbal cues [12][91][92]. The main goal of these studies was developing probabilistic models, while using existing solutions for recording the underlying parameters: interpersonal distances and relative body orientation were tracked with a precision <1mm and <1° using camera infrared beacon-based system, in the settings with cameras mounted on the floor and ceiling and beacons attached on the body [12]; speech activity was detected with the resolution of 50ms relying on the manual annotations from audio data recorded from MP3 player worn around the neck [91]. The authors demonstrated that using the abovementioned parameters was sufficient for detecting social interactions, however given an extremely high accuracy of collected data.

The previous chapters presented mobile modalities for collecting data relevant to social interactions and this chapter evaluates whether the achieved precision of detected parameters is sufficient to identify existing social interactions. Firstly, the rest of this chapter analyzes the relative predictive power of using inferred spatial parameters to identify co-located face-to-face social interactions. Afterwards, the potential of fusing speech activity and spatial settings detection is evaluated.

5.1. Detecting social interactions through spatial settings

5.1.1 Methodology

Detection of social interactions is based on analyzing the spatial parameters between a pair of subjects that carry mobile phones. If more than two subjects are involved in the same conversation, the method recognizes other participants by examining information for each pair of individuals involved in the social interaction. As the

number of participants increases, interpersonal distances expand and the angles become wider, thus imposing constraints on developing a single model of social interactions, regardless of the number of participants. However, these effects (such as changes in angles) are typically neglected in the literature since practical experience suggests that when there are more than four or five individuals, they frequently split up into sub-situations [12][24]. Therefore, the experiments that follow were conducted under the assumption that in real-life settings a number of individuals which actively participate social interaction is limited to five, referred to as a small-group interaction [12][24].

5.1.2 Experiments

The time frame of 10 seconds was chosen to process data as suggested by [92] in order to capture dynamic changes in social interactions while at the same time to discriminate between existing and non-existing social interactions. Therefore, interpersonal distances were estimated using a sequence of Wi-Fi RSSI values for every 10-second frame and also body orientation is averaged every 10 seconds (i.e. 10 samples). Relative body orientation of subjects was considered only if the standard deviation of the samples was less than or equal to 10 degrees for each subject (regarding the 10-second time frame), otherwise the current frame of samples was left out. This was done in order to analyze situations in which subjects held stable relative orientation, such that random body movements are removed as a source of orientation uncertainty. The threshold of 10 degrees was confirmed to be a trade-off between decreasing the standard deviation of the estimated relative body orientation (proportional to decreasing threshold) and decreasing the amount of discarded data (proportional to increasing threshold). Overall, approximately 20% - 25% of unstable orientation data was discarded. The application was installed in five phones, two HTC Desire, two HTC Desire S and one Samsung Galaxy S with synchronized clocks to ensure correct data aligning for the analysis. Focusing on small group co-located face-to-face social interactions, the experiments were performed in three types of scenarios described in the paragraphs that follow.

5.1.3 Controlled experiments

Participants, that partially knew each other, were asked to communicate for an amount of time of their choosing, while carrying the mobile at a known place. The first trial involved 6 participants (4 males, 2 females, age: 31±4 years) that were talking to each other, maximum four at a time, at 14 randomly selected locations, including 12 indoor and 2 outdoor environments. The duration of these interactions was 5.6±3.8 minutes. The second trial was conducted in a meeting room and it consisted of two 15-minute sessions in each involving 4 people (6 males, 2 females, age: 29±4 years) who were let to communicate freely as they wanted. This experimental trial resulted in 1300 pairs of relative body orientation and interpersonal distance (α, d) with a time frame window of 10 seconds for processing and averaging data. Figure 5.1 shows the distribution of the relative orientations and the interpersonal distances (α, d) in which the red rectangle indicates mean± standard deviation: α: 178° ± 25°, d: 1.6 m ± 0.5 m. The interpersonal distances mostly belonged to the social space (called also socio-consultive zone [38]).

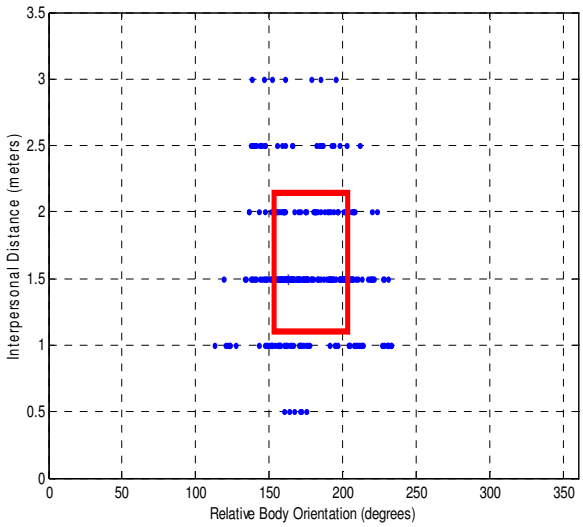


Figure 5.1: Controlled experiments

5.1.4 Break-room settings

The break room is the place where employees in the research center, which was a test-bed, typically socialize. This created the opportunity to monitor social interaction in a natural setting. When people were coming to the break room, they were asked to place the phone in a case attached on the right hip and to continue their interaction. Overall, there were 15 interactions recorded of duration 6.2±3.5 minutes that

included 24 different people. This experimental trial resulted in 1300 orientation, distance (α, d) pairs. The distribution of (α, d) is presented in Figure 5.2 (α: 177° ± 45°, d: 1.1 m ± 0.3 m). The inferred interpersonal distances belong both to personal and social space.

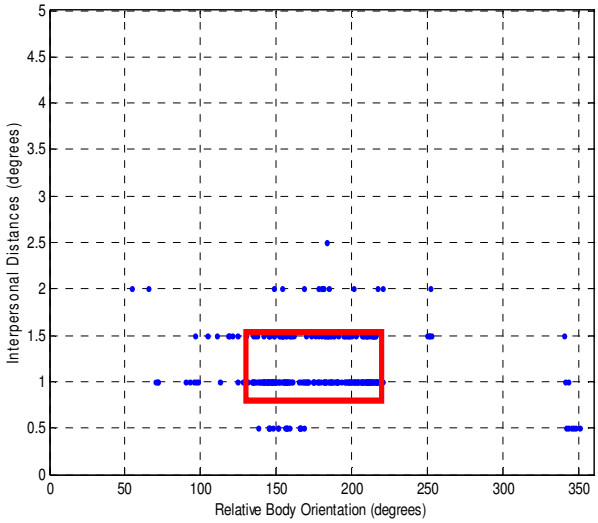


Figure 5.2: Break-room settings

5.1.5 Continuous monitoring

Aiming to analyze social interactions in continuous settings, the third trial of experiments was performed during working time for one week i.e. 5 working days, involving 5 colleagues that share the same office. They were asked to provide a label whenever social interactions occurred outside of the office, through a button press on the phone. Overall, during one week of measurements there were 9 social interactions labeled which involved either all of 5 participants or their subset. The locations were random in the building with duration of 6.4±8.1 minutes. The distribution of (α, d) for this set of experiments is presented in Figure 5.3 (α: 186° ± 45°, d: 1.5 m ± 0.6 m). A part of the inferred interpersonal distances is related to personal space while most of them belong to the social space. However, the fact that one week of measurements resulted in 9 labeled conversations was questioning; afterwards, it turned out that the participants did not label several social encounters which they clarified at the end of experiments. Thus, the next continuous experimental trial (which is analyzed later in this chapter) involved a human observer that was manually labeling ground-truth.

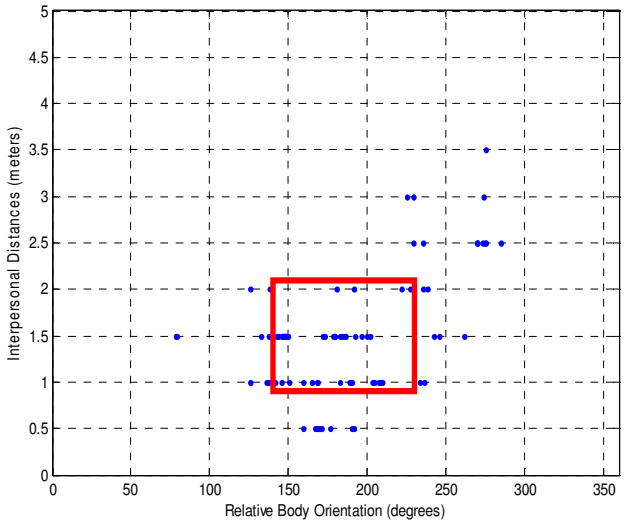


Figure 5.3: Continous experiments

Note that the absolute measures of interpersonal distances and relative body orientations as well as the comparison between the inferred distances in the three scenarios through the study of proxemics should be considered only illustratively considering the distance estimation precision (provided in Chapter 3).

5.1.6 Collecting data related to non-existing social interaction

In order to assess the potential of using spatial (α, d) parameters to distinguish existing and non-existing social interactions, it was necessary to create also a solid corpus of the pairs that do not correspond to social interactions. Four subjects that attended a fair called "Researchers Night" were monitored while being asked to report any social encounter among them. Measurements from one-hour period in which they reported no social interactions was extracted as a suitable data set containing overall 1400 (α, d) pairs for creating non-existing social interaction corpus; being at the stand implied their constant proximity and random relative body orientations (while sitting/standing/moving) - Figure 5.4. In addition, there were added measurements from previously described controlled experimental settings that included subjects that were in concurrent social interactions and in a close proximity (all social encounters occurred within 5x5m space). Certainly, there are a number of scenarios which include subjects that are not interacting while being in close vicinity, thus making it challenging to classify between occurrence and non-occurrence of social interaction. The example of one such a case involves colleagues that share the office, sitting at their

desks across each other at a short distance, but not having a conversation. This case is evaluated in the continuous experimental trial, presented in the Section 5.2.

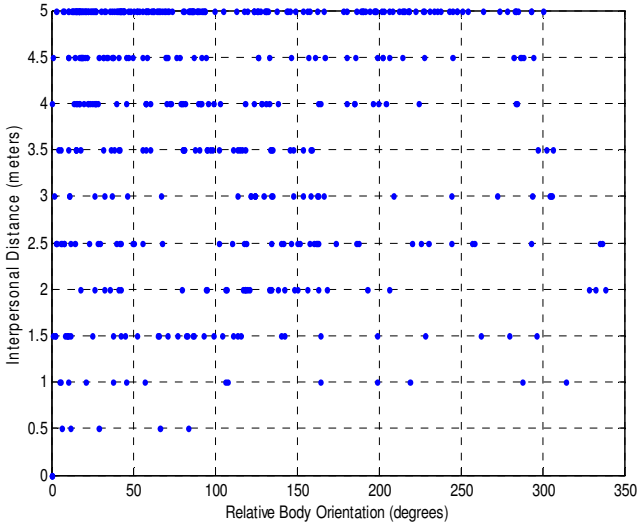


Figure 5.4 (α, d) pairs corresponding to situations without social interactions taking place

5.1.7 Data analysis

All the three experimental trials related to existing social interactions resulted in approximately 14 hours of sensor data. Figure 5.1, Figure 5.2 and Figure 5.3 indicate that conversations regarding pairs of subjects were centralized around 180° – the relative body orientation that corresponds to a perfect face-to-face position. Wider range of relative orientations was perceived in the cases of break room and continuous monitoring settings (both with SD of 45°) in comparison to the controlled experiments (with SD of 25°). This may reflect the fact that participants were more relaxed and held less steady orientation when they were participating in break room social interactions, in comparison to social interactions where participants were instructed to communicate. It can be seen that participants were mostly having shorter interpersonal distances in the break room and continuous monitoring, which were both natural setting. How these spatial parameters can be used to detect the social context (perceived by subjects as formal or informal) is analyzed in Chapter 6.

Table 5.1 presents the results of the classification between existing (denoted as SI) versus non existing social interaction cases (denoted as NonSI) represented with a feature-vector (α, d) by applying Linear Classification and Naïve Bayes with KDE techniques. The classification performance was evaluated using 10-fold cross validation.

Table 5.1: Existing versus non-existing social interactions classification (feature vector: (α, d))

	Naïve Bayes (KDE)		Linear Classifier	
	SI/ NonSI		SI/ NonSI	
SI	79%	21%	73%	27%
NonSI	24%	76%	12%	88%

The results demonstrate the accuracy of 79% in detecting social interactions based on interpersonal distance and relative body orientation. Naïve Bayes with KDE performed slightly better in identifying social interaction pairs while Linear Classifier provided lower rate of false positives. A contributing factor to a relatively high accuracy is also a simple method of taking out of the consideration (α, d) pairs corresponding to the situations in which subjects did not hold a stable relative body orientation, thus eliminating the source of uncertainty created in most cases by random body movements. However, instead of using the standard deviation (SD) of relative body orientation for identifying *unstable* (α, d) pairs, it can be also used as a classification feature that can be considered as an index of holding stable relative position of participants in a social interaction. SD of relative body orientation (denoted with σ) was also calculated for each 10-second frame (i.e. for 10 samples) and combined with distance d and averaged relative body orientation α, constituting 2-feature vector (σ, d) and 3-feature vector (σ, α, d). Table 5.2 shows the results of 10-fold cross validation.

Table 5.2: Existing versus non-existing social interaction classification, feature vectors (σ, d) and (σ, α, d).

	(σ, d)				(σ, α, d)			
	Naïve Bayes (KDE)		Linear Classifier		Naïve Bayes (KDE)		Linear Classifier	
	SI/ NonSI		SI/ NonSI		SI/ NonSI		SI/ NonSI	
SI	89%	11%	76%	24%	93%	7%	90%	10%
NonSI	31%	69%	29%	71%	26%	74%	19%	81%

The combination of interpersonal distance and SD of relative body orientation provided higher accuracy in comparison to the previous case of using relative body orientation angle (Table 5.1). This may be due to the fact that feature-vector (σ, d) does not discriminate classes based on the absolute angle between body orientations in social interactions thus allowing for more situations to be included in the model in

comparison to feature-vector (α, d). As expected, this resulted in a higher rate of false positives that occurred mostly when subjects were in a close proximity, having a stable body orientations but not interacting (for instance, sitting or being in concurrent social interactions). The highest accuracy was achieved using 3-feature vector (σ, α, d) that resulted in 93% of successfully classified vectors corresponding to social interactions and 26% of false positives (Table 5.2).

The results demonstrate that the accuracy of estimating interpersonal distances and relative body orientations achieved with mobile phone sensing was sufficiently discriminative to identify social interactions on a small spatio-temporal scale. Note that the position of the phone does not affect SD of relative body orientation thus the model based on 2-feature vector (σ, d) does not require users to carry the phone on a pre-defined/known position on the body (or using algorithms for estimating the phone position [80]).

The performance of detecting social interactions in more challenging conditions which is continuous monitoring of co-located subjects is analyzed in the following section using the two proposed modalities – speech activity and spatial settings detection.

5.2. Detecting social interactions using two-modal sensing

Four subjects that share the same office (3 males and 1 female, 29.0±1.4 years) were recruited for 7 working days. Each day, they were carrying the mobile phone at a known and fixed position on the body, typically between 11h and 17h. Accelerometer (produced by Shimmer [88]) was attached on the chest using an elastic band which was comfortable for all of the subjects except one that asked to put the sensor over a t-shirt. Overall, there were 40 hours of measurements, resulting in 452 hours of sensor data, 113.0±20.4 hours per person. In order to avoid recording (audio/video) or inquiring participants to label social interactions, ground-truth was annotated by a human observer, a colleague that shares the office with the participants, hence minimizing intrusion in typical daily routines of monitored workers. The observer manually noted each participant's speech activity and ongoing social interactions while marking all the periods in which the notes were not reliable (usually due to the lack of his pres-

ence). Labeling structure of ground-truth was divided in the two categories: 1) participation in social interactions and speech activity (present/not present) for each participant annotated every minute (52% of experimental data), 2) the existence of an ongoing social interaction, without a minute-by-minute description, including identification of participants and duration of interaction (27% of data). The rest of the data (21% of overall measurements) were lacking labels. The second category of labeling ground-truth corresponded mostly to longer discussions, conversations during lunch or coffee breaks and similar occasions during which it was cumbersome for the observer to take precise notes. The annotations were taken also for social interactions with non-monitored subjects.

Table 5.3: Speech activity detection accuracy

	Subject 1	Subject 2	Subject 3	Subject 4
	SI/Non-SI	SI/Non-SI	SI/Non-SI	SI/Non-SI
SI	67%/33%	77%/23%	73%/27%	73%/27%
Non-SI	18%/82%	21%/79%	29%/71%	25%/75%

The accelerometer data was processed providing a binary result (1/0) for every 10-second frame that indicated speech activity or not. Table 5.3 shows the results for each participant separately, analyzing only the portion of the data that was precisely annotated (minute-by-minute) – overall 30.4±9.5 hours per person. According to the structure of annotations, true positives denote the percentage of speech-labeled minutes in which speech activity was detected in at least one 10-second frame within that minute while false positives represent the percentage of minutes in which the speech was detected but not annotated in the observer's notes. It can be seen that the accuracy is slightly worse than in the case of shorter experimental trials (chapter 4). This was expected considering that due to daily activities the elastic band can move thus causing slipping or detaching sensor from the skin surface. However, the used prototype was an improvised elastic band (not purpose-manufactured) and an accelerometer that was not designed to be stuck to the skin (Figure 5.5). The lowest accuracy corresponds to the participant that was wearing the band over a t-shirt (Subject 1 - Table 5.3) that may have inhibited the chest vibrations detection.

Figure 5.5: Shimmer accelerometer on the elastic band

The resolution of 10 seconds for speech activity status was not sufficient for detecting turn-taking patterns [92] in order to identify conversation between two individuals while decreasing the frame length for processing accelerometer data resulted in speech recognition accuracy degradation. Therefore, by relying only on speech detection it was not possible to differentiate if two or more subjects have a conversation or participate in concurrent social interactions. Rather, speech activity detection was used complementary with spatial settings recognition which is described in the paragraphs that follow.

Regarding spatial settings, a classification model was built for social interactions using their occurrences from the experiments. In the test phase the whole day to which the tested social interaction belonged was excluded, hence applying a 7-fold cross-validation according to 7 days of measurements. The model for non-existing social interactions was also built and tested in the same way while excluding from the model the situations in which subjects were sitting at their desks. The reason is that due to the office layout (Figure 5.6) where the experiment took place, the corresponding feature-vectors labeled as non-social interaction would have yielded a large number of false positives. However, since the situations in which subjects were sitting at their desks corresponded to the periods annotated as non-existing social interactions, they were included in the test phase with all other situations in which no interaction was reported.

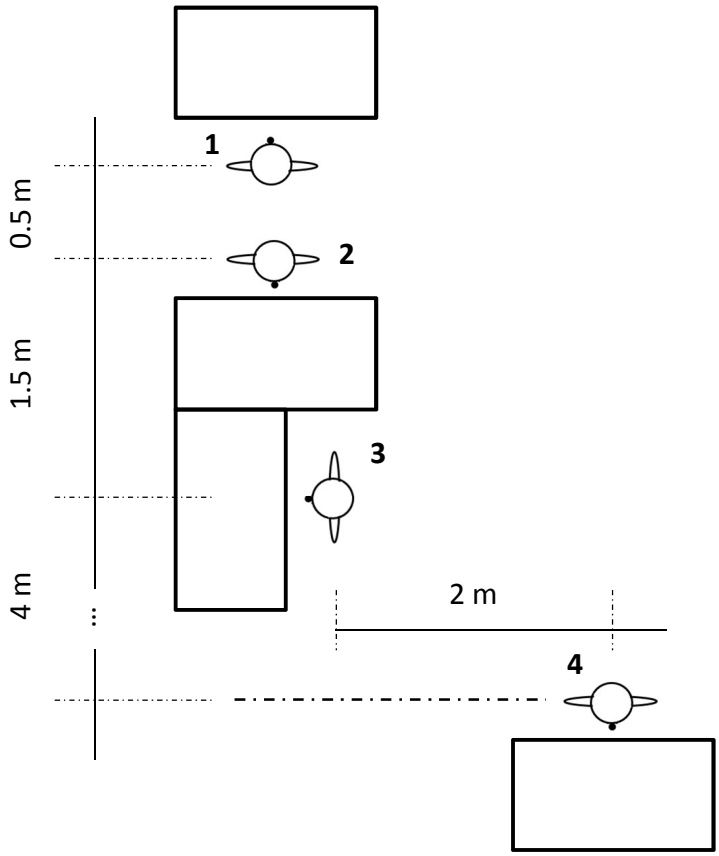


Figure 5.6: Office layout

Overall, there were 6 hours of social interactions that, according to the annotations, occurred at several locations including office (taking into account all social encounters in the office except previously mentioned "desk-to-desk" conversations), break room, meeting room, and corridors and involving two, three or four monitored workers at a time. On the other hand, 25 hours of non-existing social interactions data were annotated regardless of the location. Approximately 13% of the overall annotated data was discarded due to sporadically missing samples either from compass sensor or Wi-Fi RSSI. The experiments resulted in 1872 and 7420 feature vectors corresponding to existing and non-existing social interactions respectively.

The results are presented in Table 5.4 when Naïve Bayes classification and (σ, α, d) model were applied, which previously provided the highest accuracy. 89% of true positives and 11% of false negatives in detecting existing social interactions presents the accuracy across all the pairs of subjects. Since no major differences were witnessed for different pairs of subjects regarding true positives, the accumulated accuracy is reported. Although at an initial look (Table 5.4) the fusion of the two modal-

ities yielded no improvement in the rate of true positives, speech activity detection was used to confirm occurrence of social interactions through the presence of voice of the participants. Otherwise, if no speech was detected, even though spatial settings suggested an occurrence of social interaction, the event was categorized as a non-existing social interaction. This strategy particularly improved the identification of non-existing social interaction when relying solely on spatial settings, which resulted in a very high rate of false positives for subjects pair 1 and 2 and pair 2 and 3 (Figure 5.6). A higher rate of false positives was mainly due to the small distance and fixed body orientation of subjects in the office. Particularly for the pair of subjects 1 and 2, the feature vector of the interpersonal distance, relative body orientation and its stability (σ, α, d), was similar when sitting in the office to the feature vector describing face-to-face social interaction. This is due to the limitation of the proposed spatial settings detection system which has difficulty discriminating between a back-to-back and face-to-face position of subjects. However, in the current experimental scenario, the fusion of the two sensing modalities significantly improved the overall results. A portion of false positives was resolved by checking speech activity status whenever spatial settings analysis indicated an existing social interaction. In these cases, if the speech activity was not recognized for both subjects during an arbitrarily selected time frame of 5 minutes, the system indicated non-existing social interaction. For the pair of subjects 1 and 2, the results for false positive showed the drop from 76% to 34% and for the pair 2 and 3 from 39% to 29%. The fusion of the two sensing modalities also contributed considerably in resolving false positives for other pairs of subjects that occurred mostly due to their random daily movements. It should be mentioned that this method for resolving false positives did not negatively affect the recognition rate for existing social interactions.

Table 5.4: Results of two-modal sensing of social interactions

	Spatial		Spatial + Speech	
	SI/ NonSI		SI/ NonSI	
SI	89%	11%	89%	11%
NonSI				
Sub 1&2	76%	24%	34%	66%
Sub 1&3	19%	81%	11%	89%
Sub 1&4	17%	83%	15%	85%
Sub 2&3	39%	61%	29%	71%
Sub 2&4	15%	85%	15%	85%
Sub 3&4	17%	83%	14%	86%

5.3. Summary

This chapter presented the experimental results which evaluated the possibilities for detecting social interactions relying on inferred spatial settings between subjects and speech activity status. As evidenced from the experiments, these two modalities provide complementary information about social interactions with a sufficiently high precision to indicate the occurrence of social encounters.

Using solely speech activity detection does not suffice for a reliable detection of social interaction, yet in this way the overall amount of speech during a certain interval can be estimated, thus reflecting the participation in verbal social interaction. On the other hand, in certain scenarios, combining interpersonal distances and relative body orientations demonstrated the high predictive power despite concurrent interactions occurring in a close proximity. However, situations in which subjects hold, for a prolonged period, the position which may indicate a conversation albeit not interacting can result in a higher rate of false positives. This can be resolved by including the knowledge of speech activity status, which is used to confirm or reject the occurrence of a social interaction suggested by inferred spatial settings. Therefore, the most accurate interaction detection was achieved by relying on the fusion of inferred speech activity status and spatial settings parameters.

Chapter 6

6. Recognizing type of social interactions

In 1995 Savage [93] described the future in which 2% of the working population will work on the land, 10% will work in industry and the rest will be knowledge workers. Although such future has not come true yet, the trends in previous years indicate that Savage's predictions were not random – according to the statistics [94], knowledge workers already constitute 70% of the labor force in the US. While the productivity of manual worker has been thoroughly investigated and resulted in a variety of strategies for its improvement, increasing knowledge workers' productivity is more complex and still little is known about the underlying principles. Mc. Dermott [95] estimated that 38% of time knowledge workers spend searching for information which is the fact that motivates a promising avenue for investigation – how to improve the ways for exchanging and distributing information in order to increase the productivity [96][97]. Various studies investigated the methods for improving communication channels to enable more efficient knowledge exchange between employees. Most of the outcomes suggested the promotion of informal type of communications [98] which was demonstrated to play a crucial role in maintaining work and for the overall success of a company. Yet, there are several studies showing the opposite i.e. arguing for formal interactions as the way for an efficient knowledge transfer [99]. In the attempt to emphasize both, Krackhardt and Hanson described: "If the formal organization is the skeleton of a company, the informal is the central nervous system driving the collective thought processes, actions, and reactions of its business units" [100]. However, there is a general consensus that improving communication channels require a deeper understanding of both formal and informal types of interactions [98] [99] [41].

Despite the fact that enterprises were increasingly investing in projects designed to improve knowledge management, the lack of theoretical findings and proven approaches still limits significant movements in this field [101][97]. The difficulty in

monitoring and measuring informal/formal networks was identified to be a key challenge towards making substantial steps in the efficient information transfer and consequently for increasing productivity in knowledge-driven communities [41]. Similarly to the standard methods applied in a number of other domains, formal/informal communication networks were being investigated in the past, relying on the help of observers or on self-reports collected through interviewing participants of the study. Such methods of reconstructing data were error prone [102] and even lead to contradictory results in this area [41]. The reason mostly lies in the fact that periodical surveys fail in capturing the temporal aspects i.e. neglect individual social interactions [103] while social networks in companies are typically characterized by dynamic changes.

With the view of overcoming the limitations of self-reporting methods, the use of automated data collection for allowing insight into formal/informal structures would essentially contribute in acquiring new theoretical findings on knowledge transfer [41]. Nonetheless, the problem has been addressed only by few studies in ubiquitous computing community. The analysis of social networks was facilitated by using sociometer [8], [42], [70] and also by relying on mobile phones [104], however these studies were mostly intended to map the structure of networks by inferring who and when interacted. Since the content of audio was not analyzed due to privacy concerns, the type of individual social interactions was not revealed. Along this line, in the recent study, Do and Gatica-Perez [64] recognized the types of social interactions by analyzing continuously sampled Bluetooth data. In order to infer the type of interactions the authors relied on longitudinal data analysis thus not focusing on temporal aspects (for instance, the two same colleagues can be engaged in one informal and another formal interaction during the same day which would not be distinguished).

This chapter evaluates the potential of using the proposed sensing modalities to indicate the type of social interaction once its occurrence is already detected on small spatio-temporal scale, as elaborated in the previous chapter. Although the term *type* of social interaction may include various connotations (such as competitive, cooperative, decision making, and other types of conversation), it is used here to denote formal or informal context. Nonverbal cues that can be extracted using the proposed system are evaluated regarding the predictive power in classifying between formal and informal

type of interaction. The goal is to provide a tool for acquiring a better insight into the contexts of individual social interactions thus to potentially support the research in social networks analysis with respect to formal/informal structures.

The results are based on the experiments that included overall 53 social interactions with the duration of 12.2 hours that occurred in natural and unconstrained settings.

6.1. Formal and informal social interaction

Most of the work done in knowledge-driven organizations requires a certain level of collaboration among workers, which implies the need for communication. Depending on the job type, it is estimated that between 25% and 70% of time workers spend in face-to-face interaction [105] which includes both formal and informal ways of communication. Despite the common understanding of the distinction between these two types of interaction, the concept of formal/informal communication lacks a clear-cut definition typically varying depending on the scientific discipline, moreover the field of the study which examines it. Considering the goal of identifying social contexts regarding knowledge workers, this work establishes the relation between formal and informal interactions referring to the field of social psychology. Referring to the social psychology literature, Kraut et al. [98] described several variables for discriminating between formal and informal communications – time scheduling, involved participants and their roles, agenda, content and language of the conversation (Figure 6.1). By analyzing the set of variables, the authors aimed to illustrate the characteristics of the two types of social interactions but also to provide the foundation for designing the questionnaire for classifying social encounters.

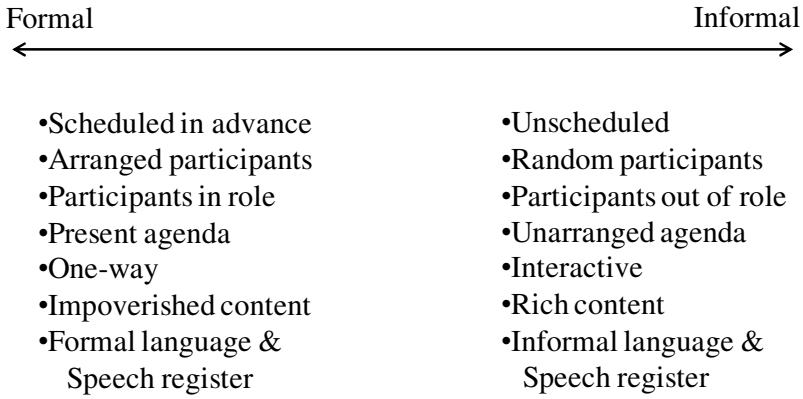


Figure 6.1: The formality dimension of communication (taken from [98])

6.2. Inferring the social context ground-truth

The questionnaire, used to infer the ground-truth (formal or informal social context) in monitored social interactions in the experiments that follow, was designed and processed according to the instructions provided by Kraut et al. [98] which categorized the context relying mostly on the degree to which the conversation was scheduled. The four categories for assessing the degree of preplanning conversation included: a) scheduled (previously scheduled/arranged interaction), b) intended (there was one initiator prompting other subject for the conversation), c) opportunistic (one participant planned to talk with another and took the advantage to have a conversation), d) spontaneous (there were no previous plans for the conversation) [98]. Each participant responded independently and the conversation was characterized with the least spontaneous answer following the order of scheduled < intended < opportunistic < spontaneous (for instance, if one reported opportunistic and another scheduled, the conversation was categorized as scheduled). However, for further analysis of monitored social interactions, the questionnaire used in the experiments related to this chapter included also demographic information, topic of conversation (work related, non-work related), frequency of the communication between participants (every day, several times a week or not regularly), period that participants knew each other (less than 3 months, between 3 months and one year, more than one year), and subjective description of the conversation (formal, informal).

6.3. Spatial and speech activity cues for informal vs. formal interaction classification

6.3.1 Speech activity cues

Several variables from the set that indicate typical differences between formal and informal interactions (Figure 6.1) correspond to speech activity characteristics of the conversation, including the level of interaction/one-way-speech, richness of the content, speech register and the degree of language formality. Since the content of conversations is not available, the level of formality, spectrum of content and speech register cannot be automatically extracted. However, the level of interaction can be estimated based on the amount of talk for each participant, information that is available from the proposed speech activity detection modality which provides speech status with a resolution of 10 seconds. This makes the distribution of the amount of time that each participant was speaking a suitable nonverbal cue for the classification between formal and informal contexts.

Furthermore, several characteristics of the informal/formal interaction distinction (such as roles of participants) may be reflected through speech patterns. This particularity refers to the findings that dominance and status of participants in social interactions are correlated with speech related cues including speaking energy [106], speaking length and turns [107], and interruptions in conversation [108]. In the recent work, Jayogopi [22] used technological solutions to extract a number of nonverbal cues in order to estimate status and dominance in social interactions, including: speaking energy, speaking length, number of speaking turns, turn duration, number of successful and unsuccessful interruptions, and several derived cues. The study demonstrated that the aforementioned cues contribute to detection of the most dominant person, the status of an individual but also to group conversational context identification (regarding “competitive vs. cooperative” and “brainstorming vs. decision making” classifications).

However, speech energy estimation relies on the information about the speech loudness while, as previously discussed, identifying turns in speech requires the resolution in detection significantly higher than 10 seconds [92]. In this regard, Jayogopi [22] used the four close-talk microphones attached to participants and analyzed audio data each 40 ms with a 10 ms time shift. Thus, all the cues related to interruptions and

turns cannot be extracted from conversations using the proposed accelerometer-based system for speech detection. Although it remains on the speculative level that these cues can contribute to the formal/informal context classification, it may be a promising area for future investigation, not yet being addressed by the current literature.

However, speaking length can be effectively recognized using the accelerometer-based system affording the extraction of the two relevant cues, namely speaking length index and speaking length distribution index (both cues referring to the group behavior) [22].

Speaking length index was calculated as the sum of total amount of time that each person spent speaking divided by the overall interaction duration. Speaking length distribution index was calculated following the algorithm which was reported by Jayogodi [22]:

1) Compose the vector A representing the levels of participation for each subject in the conversation with respect to the others, $\frac{t(i)}{\sum_n t(i)}$,where n is the number of subjects and t(i) represents an amount of time that i-th subject was speaking during the overall duration of social interaction.

2) Using Bhattacharyya distance [109], compare vector A with the uniform vector which is of the same dimension n being constituted of values $\frac{1}{n}$.

This method yields a value between 0 and 1 for each social interaction, where 0 would correspond to the social interaction in which all participants have spoken an equal amount of time while 1 corresponds to a conversation in which solely one participant was talking. Note that speaking length distribution index reflects the above discussed variable "one-way/interactive" (Figure 6.1), which characterizes the difference between formal and informal context.

6.3.2 Spatial cues

The main postulates of the proxemics study presented in Chapter 3 suggest that people unconsciously organize the space around them, corresponding to different degrees of intimacy. It is even intuitively known that having a chat with a close friend and talking to the boss differ in spatial settings conventions i.e. that setting interpersonal distances is affected by level of formality in social interaction. Furthermore, ac-

cording to social psychology, the formality is bounded with roles and hierarchies among participants [98] (in Figure 6.1 illustrated with the variable "participants in/out of role) which is further mirrored in spatial arrangements. The matching between social relations and the spatial formations in social interactions was recently investigated using computer vision system for estimating distances between subjects, confirming strong positive correlation [38]. Therefore, the choice of interpersonal distance was straightforward in the attempt to classify between formal and informal social context.

Regarding spatial settings detection, the proposed system provides measures of relative body orientation and its standard deviation (as an index of stable relative body position between participants) which demonstrated high predictive power of detecting social encounter occurrence. Social psychology literature does not directly associate body orientations and the degree of formality in conversations. However, the relative body orientation is often used in studies to describe the immediacy of interaction, subject's attitude or similar phenomena in social interactions [79]. Therefore, it is hypothesized that the body orientation related cues (namely relative body orientation and its standard deviation) might also correlate with the level of formality thus being selected as suitable cues aiming to formal vs. informal interaction classification.

6.3.3 Overview of the classification problem - temporal and cumulative cues

It can be noted that selected cues for formal versus informal interaction classification are of different nature that is discussed in the paragraphs that follow.

Interpersonal distance, relative body orientation and standard deviation of relative body orientation are captured in temporal terms, every 10 seconds during an ongoing social interaction, and will be referred to as temporal cues. These temporal cues are calculated for each pair of subjects that participate in the same social interaction.

In contrast, speaking length and speaking length distribution indices result in one value that characterizes the completed social interaction, which will be referred to as cumulative cues. Unlike temporal cues that refer to pairs of subjects in conversation, cumulative cues describe an entire group behavior during a social encounter. In addition, location and duration of conversation, attributes that can be assigned to each concluded social interaction, are combined with the cumulative cues attempting to improve the classification accuracy. Location is expressed as an index representing the

probability of informal social interaction occurrence calculated solely based on the experience from the experiments (for instance, if 4 formal and 6 informal social interactions occurred in a building hall, this location is assigned with a value of 0.6). Note that expressing location index as the probability of a formal instead of informal conversation event would be equivalent given that the two occurrences are mutually exclusive. Location was automatically detected using the mobile phone, however in most cases it was known in the experiments; the elaboration on the possibilities of using mobile phone for indoor positioning is presented in [78]. Duration refers to a number of minutes from initiating until concluding the conversation. Certainly, duration and location of formal and informal social interactions are strongly dependent on the test-bed being affected by various parameters typically specific for a certain workplace, building layout or workers' routines (while other selected cues reflect general phenomena). The goal of including these two attributes in the analysis of formal versus informal context classification was to investigate whether a heuristic-based approach can contribute to the classification, bearing in mind that its applicability remains limited to an individual test-bed.

The summary of the attributes (highlighted with grey color) and the temporal/cumulative cues which are evaluated regarding formal and informal social interaction classification is given in Table 6.1 with corresponding denotations.

Table 6.1: Summarized cues intended for formal vs. informal classification

Interpersonal distance	d
Relative body orientation	α
SD of relative body orientation	σ
Speaking length index	***SLI***
Speaking length distribution index	***SLDI***
Duration of social encounter	***DUR***
Location index	***LOCIN***

6.4. Experimental setup and meeting data

The experiments were conducted in several locations, including three meeting rooms, three offices, three coffee rooms, two balconies and an entrance hall with dimensions that did not physically confine subjects thus not affecting interpersonal distances. At randomly determined times, face-to-face interactions that were about to oc-

cur or were already initiated, were interrupted by explaining to subjects that the investigation is on social interactions phenomena which does not require recording data. Afterwards, they were provided with the equipment for monitoring (accelerometer for speech activity detection and smart phones that were sampling orientation and broadcasting/receiving Wi-Fi signal). In most cases, subjects were accepting the participation in experiments, however often refusing accelerometer due to inconveniencies related to time required for mounting it on the chests. They were given a case to carry the phone on the right hip thus the position of the phone with respect to the body was directly known in order to calculate the relative body orientation. Once the social interactions ended, participants were asked to fill out a short check-box questionnaire that was previously described in section 6.2 in order to infer whether the conversation was formal or informal. Overall, there were 30 face-to-face interactions collected, 21 informal and 9 formal, which included participation of 50 subjects (33 males/17 females, with an age of 32.7±6.6 years) resulting in 11.2 hours of sensor data. Wi-Fi and orientation were sampled with 1Hz and interpersonal distance and the relative body orientation were estimated for each time frame of 10 seconds. Exact duration was captured in 16 informal (duration of 9±5 minutes) and 9 formal (duration 21±9 minutes) provided that subjects were asked to participate the experiment before the conversation. Only 2 formal and 4 informal interactions were monitored including sensing modalities, the accelerometer and mobile phones. However, during the continuous experiments described in Section 5.2 which involved both speech and spatial settings detection systems, participants were asked to report every scheduled meeting they encountered with monitored participants which resulted in 7 formal and 16 informal meetings with duration of (25±8 minutes) and (8±7 minutes) respectively.

Therefore, there were overall 54 monitored meetings (37 informal and 17 formal), out of which 20 informal and 9 informal were monitored with both sensing modalities (speech status and spatial settings recognition) and the rest with mobile phones phone (providing spatial arrangement detection). This resulted in approximately 450 hours of sensor data related only to time spent in conversation.

Table 6.2 presents the results from questionnaires collected after every social encounter involving each participant to fill it out independently. In several cases when answers were not in concordance, single social interaction was assigned with the least

reported value in terms of selecting the smallest reported frequency of communication, the smallest reported amount of time that subjects knew each other, the most formal reported context and the least formal topic of conversation. *Opportunistic* and *Spontaneous* meetings were always subjectively described as informal while scheduled ones were described as formal conversation, being in accordance with the distinctions between formal and informal context provided in [98]. Although intended meetings were suggested to be categorized as informal, the participants were mostly reporting formal contexts as the subjective description. In these cases, the topic determined the formality assigning the attribute *formal* given that the subjective description was formal and the topic was work-related, otherwise the conversation was categorized as informal. Overall, 70.3% of meetings were informal while 29.7% occurred in a formal context. It can be noted that formal way of conversation was not associated with the frequency of interaction and the period which subjects knew each. In other words, subjects were interacting formally regardless of how much they were familiar with each other. On the other hand, informal interactions were mostly occurring among subjects that knew each other better.

Table 6.2: Social context ground-truth

	% of overall no. of meetings	Topic		Frequency of communication		Time of being acquaintances		Subjective description	
		work-related	other topics	regularly	non-regularly	<6months	>6months	formal	informal
Scheduled	13%	100%	0%	25%	75%	25%	75%	100%	0%
Intended	20%	83%	17%	50%	50%	33%	67%	83%	17%
Opportunistic	17%	60%	40%	20%	80%	20%	80%	0%	100%
Spontaneous	50%	47%	53%	7%	93%	13%	87%	0%	100%

In continuous experimental trial described in Section 5.2 subjects were asked to annotate each scheduled meeting while one part of informal conversations were reconstructed from observer's notes. In this way, only meetings with the ascertain ground-truth were processed while social encounters without reliable annotations were not considered for analyzing formal and informal contexts.

Despite a limited number of trials, the results appear to be consistent with the literature indicating that knowledge workers mostly communicate in informal way while using that communication channel also for discussing work-related matters [98][41][105]. However, the goal of this study was to analyze the predictive power of

the selected cues, extracted using the proposed mobile sensing modalities, for formal vs. informal interaction classification.

6.5. Formal versus informal interaction classification based on cumulative cues

Speaking activity cues were extracted using accelerometer-based approach (Chapter 4) while the duration of meetings was directly available (Section 6.4). Location with the granularity reduced to the room level was detected applying Wi-Fi fingerprinting method (fingerprints were previously captured in the locations of interest) and in more than 90% of cases matched the manual annotations.

The predictive power of the cumulative cues and attributes in classifying the social contexts is assessed through the cross validation and the results are presented in Table 6.3. The accuracy corresponds to the percentage of social interaction occurrences correctly classified as formal or informal according to the ground-truth.

Table 6.3: Formal vs. Informal classification (cumulative cues)

Feature	Naïve Bayes (KDE)	SVM
SLI	55%	55%
SLDI	62%	66%
SLI+SLDI	66%	66%
SLI+DUR	66%	72%
SLI+LOC	66%	66%
SLDI+DUR	72%	76%
SLDI+LOC	72%	72%
SLI+LOC+DUR	69%	69%
SLDI+LOC+DUR	76%	79%

According to the results, Speaking Length Index (SLI) was not shown to be indicative for discriminating between formal and informal context providing the accuracy of 55%. It can be noted that a random guess would provide the accuracy of 50%. Combining SLI with the location or duration of meetings, the accuracy increased by up to 10%. As expected considering social psychology literature, Speaking Length Distribution Index (SLDI) that reflects the variable “one-way/interactive” (Figure 6.1) performed better, showing moderate accuracy in discriminating between the two types

of conversations. This suggests that in informal social interactions participants spent more equal amount of time talking than in the case of formal context. The empirical distribution of SLDI is presented in Figure 6.2 for the two types of interactions. When SLDI was combined with the location and/or duration of social interactions, the accuracy was increased up to 79%, while the fusion of SLI and SLDI did not improve the results.

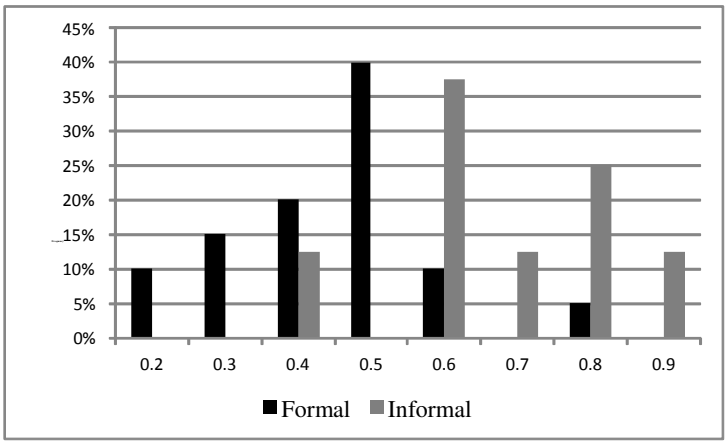


Figure 6.2: SLDI distribution

When fused with the selected cumulative cues, location and duration of social interactions augmented the classification accuracy. The two attributes can be automatically extracted using the proposed mobile modalities but unlike the selected cues they require building the heuristics for a specific test-bed that limits their application.

Speaking Length Index which represents the proportion of time that all the participants together used during the overall duration of a social encounter was not shown to be discriminative. Speaking Length Distribution Index, which is reported in the literature to be very effective for estimating the most dominant person, demonstrated a moderate accuracy in classifying between formal and informal context – being successful in 66% of cases. The best achieved accuracy of 79% was in the case of combining SLDI with both location and duration of social interactions, indicating that in specific applications (when the occurrence of social interaction is witnessed) speaking activity related cues extracted by using solely an off-the-shelf accelerometer can distinguish between formal and informal interaction context.

6.6. Formal versus informal interaction classification based on temporal cues

Whereas cumulative cues refer to the group characteristics of a finished conversation, temporal cues are variable for each pair of subjects during an ongoing social interaction. Prior to evaluating the accuracy in discriminating between formal and informal meetings, the predictive power of temporal cues for the classification is examined by analyzing their distributions.

6.6.1 Interpersonal Distances

Figure 6.3 shows the histogram of interpersonal distances, plotted for each time frame of 10 seconds recorded during formal and informal communications. According to the study of proxemics [19], interpersonal distances detected in informal communications mostly belong to the Personal Space having the mean value of 0.8 m. In formal communications, the results show distances that correspond to both Personal and Social Space, with the mean value on the border of these two zones, at 1.3 meters. These absolute measures should be taken only illustratively considering the distance estimation precision (provided in Chapter 3). However, both distributions in Figure 6.3 were acquired using the same system thus embedding the same median error in estimated distances. Therefore, whereas the absolute measures cannot be reliably claimed with a precision less than 50 cm due to the system's accuracy, the relative difference between interpersonal distances may provide more reliable estimate on the actual phenomenon. Furthermore, the results demonstrate that the distinction between formal and informal social interactions is reflected in interpersonal distances estimated using mobile phone sensing despite the median accuracy of 50 cm. Due to the considerable intersection of values related to interpersonal distances in formal and informal social interactions, relying solely on this temporal cue to distinguish the two types of social interactions would not suffice. Rather, it is investigated if interpersonal distance can be combined with other selected temporal cues to distinguish different social contexts.

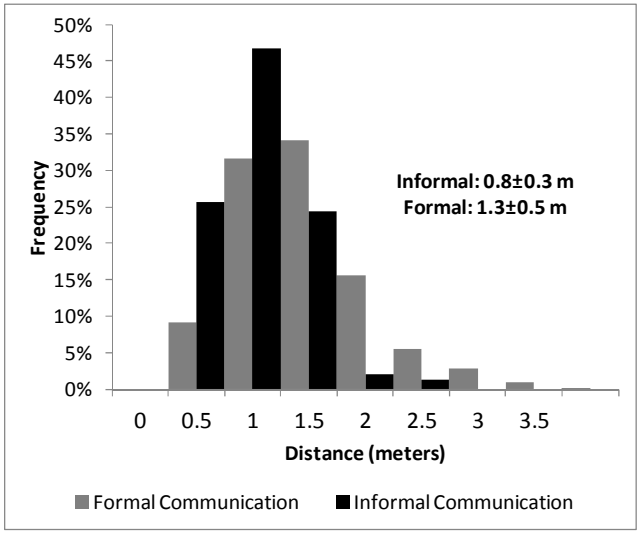


Figure 6.3: Interpersonal distances in formal/informal social interactions

6.6.2 Relative Body Orientation

Relative body orientations in formal and informal social interactions were analyzed after discarding all 10-seconds frames with the standard deviation greater than 10 degrees in order to capture only the readings corresponding to a stable position between subjects (Figure 6.4)

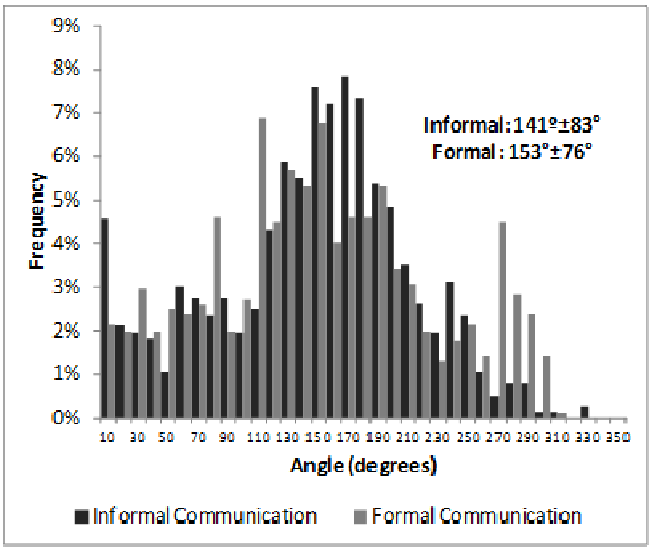


Figure 6.4: Relative body orientations in formal/informal social interactions

In both formal and informal communications the mean value of relative body orientation was between 140 and 150 degrees (180 degrees corresponds to a direct face-to-face orientation) thus not demonstrating major differences between the two types of communication. It cannot be concluded if such results pertain to the phenom-

enon of formal/informal communications or it was due to the limited accuracy in estimating relative body orientation using the compass sensor embedded in phones and approximating the angle between the body and the phone orientation. However, when recognized with mobile phone, relative body orientation did not mirror the difference between formal and informal conversational context.

6.6.3 Standard deviation of relative body orientation

Figure 6.5 shows the histograms of SD of relative body orientation for each 10-second frame during formal/informal social interactions. The results indicate that subjects were more flexible in holding the relative body orientation during informal communications than in the case of formal interactions. In formal interactions relative body orientation of subjects had a tendency to remain stable for longer periods (in contrast to informal social context), which may be due to maintaining eye contact for example, or some other external factor such as a video beam or a monitor that focused subjects' attention. Therefore, among selected temporal cues, classification between formal and informal communications is evaluated on the basis of interpersonal distance and the standard deviation of relative body orientation.

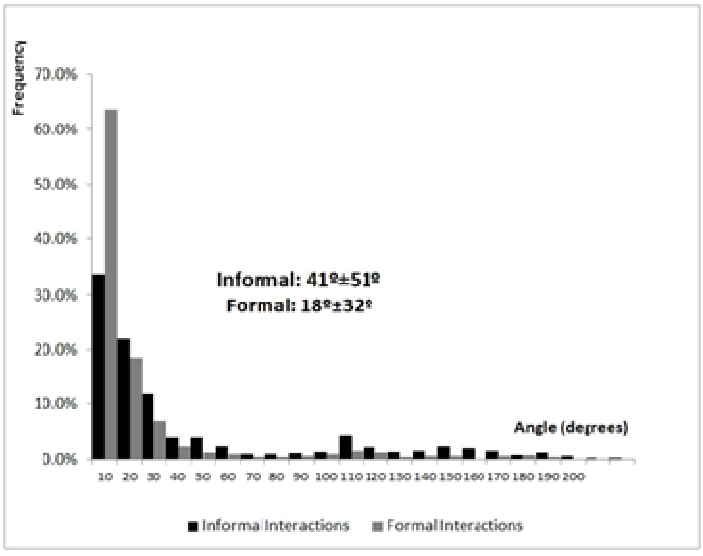
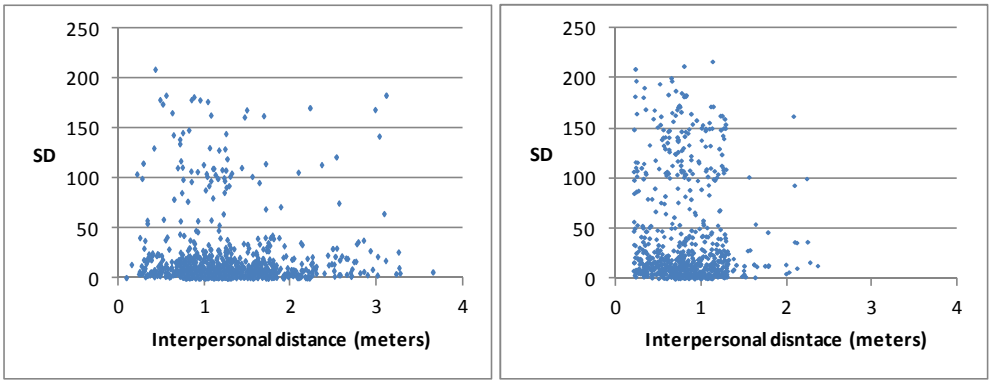


Figure 6.5: Standard deviation of relative body orientations (calculated each 10-second frame)

6.6.4 Classification results

The pairs of interpersonal distance and standard deviation of relative body orientation calculated for each 10-second time frame are plotted in Figure 6.6 separately

for formal and informal social context. The visualization of the data shows the differences between formal and informal interactions across the two temporal cues thus further indicating their discriminative potential. Please note that interpersonal distances were estimated applying GP regression for a more precise illustration (unlike in Section 5.1) of the differences between the two types of social interactions.



a)Formal interactions b) Informal Interactions

Figure 6.6: Interpersonal Distances and Standard Deviation of Relative Body orientation plotted pairwise

The classification results (Table 6.4) demonstrate that interpersonal distance and standard deviation of relative body orientation are well suited features to discover the type of face-to-face communication, providing the maximal accuracy of 81%. Furthermore, computing both parameters does not require the phone to be at a known place on the body thus affording an unobtrusive monitoring of subjects that habitually carry mobile phone.

Table 6.4: Formal vs. Informal classification (temporal cues)

Feature vector	Naïve Bayes (KDE)	SVM
(d, α)	67%	68%
(d, α, σ)	76%	78%
(d, σ)	78%	81%

6.7. Summary

This chapter investigated the possibilities of using the proposed sensing modalities for automatic classification between formal and informal types of social interac-

tions. Referring to the social psychology literature helped in identifying nonverbal cues that are meaningful and informative for interpreting the social context. The evaluation demonstrated the high predictive power of spatial settings parameters for formal versus informal classification, resulting in the accuracy up to 81%. Spatial nonverbal cues were extracted solely by using mobile phone sensing, not required to be at a known place on the body. On the other hand, extracting nonverbal cues from speech data was limited by the detection resolution of 10 s provided by the accelerometer-based speech activity detection. However, one of the selected nonverbal cues which reflects the level of interactivity of a group conversation, namely Speaking Length Distribution Index, demonstrated a moderate accuracy in classifying between formal and informal context – 66%. When combining Speaking Length Distribution Index with indices related to a location or a duration of social interaction (built through the heuristics for a specific test-bed), the accuracy increases to 79%, indicating that in specific applications (when the occurrence of social interaction is witnessed) speaking activity related cues extracted by using solely an off-the-shelf accelerometer can distinguish between formal and informal interaction context.

Monitoring social interactions has a particular application in workplace, where socialization patterns can be used to influence policies towards a more efficient work environment. Various studies investigated methods of improving communication channels to enable more efficient knowledge transfer between employees. In the recent years, social scientists have debated which type of social interaction is more productive, formal or informal. However there is a general consensus that improving communication channels used by knowledge workers requires a deeper understanding of both formal and informal types of interactions which pertains to their monitoring and assessing.

Chapter 7

7. Social interactions and emotional response

The association between social encounters and emotions has been long known. One of the most famous Shakespeare's plays, Othello, portrays characters from different backgrounds whose happiness depends mostly on social interactions. On the other hand, it holds intuitively that happy persons will more likely participate in social interactions than those who are feeling sad while talking to a close friend about personal problems can make one feel better. However, scientific evidence on the association between social relationships and psycho-physical health has been established in the late 1970s and the 1980s [110]. Since then, this domain have been attracting the attention of both social psychology and health related sciences; it was demonstrated that subjects with a low quantity of social relationships are less healthy, psychologically and physically, while manifesting higher risks for tuberculosis, accidents, and psychiatric disorders such as schizophrenia [110]. Recent studies suggested that an increased amount of social interactions can improve depressive symptoms [111][112]. In addition, individuals who maintain a certain level of social engagements are shown to be more successful in coping with stress, and in the case of the elderly, they are highly functional and independent [113]. However, while people show awareness of the general recommendations regarding physical activity and diet, they typically neglect other factors that impact wellbeing, such as social activities [113]. Therefore, social interactions become an important aspect for monitoring also regarding psycho-physical health.

The previously discussed shortcomings of gold-standard surveys for interaction data collection exhibit even additional issues when investigating psychological aspects. Firstly, self-reports are subjective and the outcome depends on the subjects' current mood that can result in an incorrect description of past activities. Secondly, individuals are prone to neglect certain parameters that influence their mood or overestimate the influence of other aspects (such as a very small effect of weather in indi-

viduals' day-to-day mood despite the commonly held conception that weather greatly affects the mood [114]). In addition to the quality of collected data, the importance of uncontrived experimental conditions is underlined through the fact that obtrusive methods and the lack of privacy can easily affect subjects' mood and consequently result in misleading data that is collected.

This chapter demonstrates the use of the proposed sensing modalities to investigate the correlations between social activity and the mood changes that are expressed through the dimensions of PA (positive affect) and NA (negative affect). The current literature in social psychology reports several studies that examined how the social activity impacts the mood during the day [115] [116] [117] [118] [119], however none relied on the automated methods for collecting data.

7.1. Methodology

One's mood may depend on a number of different factors, such as circadian rhythms [120], type of environment [121], quality of sleep [122], state of health, private problems or some other factors incomprehensible not only through direct measurement but also difficult for an individual himself/herself to identify. Therefore, it may be impossible to consider all the factors that influence the mood and provide the ultimate conclusion about the exact cause of one's state of mood. For this reason, this study follows relative changes of the mood dimensions of PA/NA rather than to focus on an absolute mood state, while assuming that interval between two mood assessments of a couple of hours (in the experimental design) is not sufficient for a significant change in *background* factors. It is hypothesized in the following investigation that these factors, such as private problems for example, are likely to be constantly present during relatively longer periods of time while, the activities within that period have pre-dominant influence on relative changes of mood. The goal of this research is to capture patterns of social activity that, in most cases, provoke similar responses in individuals' mood.

7.1.1 Monitoring social activity

As previously demonstrated, the solution presented in this thesis reliably detects the occurrence of social interaction and also a set of interaction features. Through the three pilot studies, this chapter analyzes whether the extracted parameters related to social activities correlate with the mood changes:

The first study examines the correlation between the amount of speech, which is one aspect that reflects social activity, and the mood changes. The amount of speech was estimated using the accelerometer-based method.

The second study assesses the correlation between an amount of pleasant social interactions and the mood changes. The speech activity detection and the localization system are used to isolate a set of interactions perceived from workers' point of view as pleasant.

The third study involves the combined sensing system which, in comparison to the previous two studies, identifies also the participants of detected social interactions (out of monitored subjects). This study aims to capture if social interactions between certain subjects induced consistent responses in their mood.

7.1.2 Measuring mood changes

While detecting social interaction parameters in an automatic manner, the mood in subjects was measured relying on the standard, questionnaire based method. Despite of an increasing attention that the field of automatic mood recognition has been receiving, the practical use of such methods, as a reliable alternative to standardized questionnaires, has not been demonstrated yet. Therefore, the method of assessing mood fluctuations during the day was based on EMA (Ecological Momentary Assessment) approach in order to compare retrospective and momentary mood data [123]. The EMA approach, which involves asking participants to report their psychological state multiple times a day, reduces the critical issue of retrospective recall of extended time intervals. The retrospective recall is related to cognitive and emotive limitations that bias the recall of autobiographical memory influencing subject's report by most salient events during the recall interval. The questionnaire used in this study was derived from a well-established scale – the Profile of Mood States (POMS) [124] that consists of 65 items in its standard version. However, long and repeated mood questionnaires become a burden on subjects; therefore 8 adjectives from the POMS

scale were derived, namely cheerful, sad, tensed, fatigued, energetic, relaxed, annoyed and friendly that were rated on 5-point scale (1-not at all, 2- a little, 3- moderately, 4 quite a bit, 5- extremely). The points were summed across the items related to PA and NA dimensions while the difference in scores between two sequential questionnaires was taken as a measure of relative change of subject's mood states. The questionnaires were administered three times a day, scheduled to best fit with the office workers' routines which were recruited in the experiments. Typically, the questionnaires were answered in the morning, after lunch and at the end of working day.

7.2. Speech activity and mood changes

The current literature reports several studies that examined how the social activity impacts the mood states during the day [115] [116] [117] [118] [119]. Vittengl et al. [115] and Robbins et al. [116] demonstrated that different types of social encounters provoke diverse emotional effects, while there is also an association between the overall amount of social interactions and responses in positive affect [117][118][119]. All previous referenced studies were consistent in revealing the positive relation between social events and the mood dimension of positive affect (PA), while negative affect (NA) factors were shown to be correlated either with only certain types of conversations or not associated with social activity at all.

This study investigates the correlation between self-reported mood changes and the overall amount of speech within a certain interval which reflects participation in verbal social interactions.

7.2.1 Experiments

In order to estimate the amount of speech activity within a certain period, the accelerometer produced by Shimmer [88], was attached on the chest, continuously sampling and storing the data. Applying the model described in chapter 4, each 10-second time frame of the acquired data was separately queried and classified according to the presence of speech. Afterwards, for each interval of interest, the number of minutes in which at least one 10-second frame indicated speech status was calculated thus providing an aggregated number of minutes in which subjects were speaking. Overall 10 knowledge workers (7 males, 3 females) were recruited during one work-

ing week (5 working days). The characteristics of the sample are presented in Table 7.1.

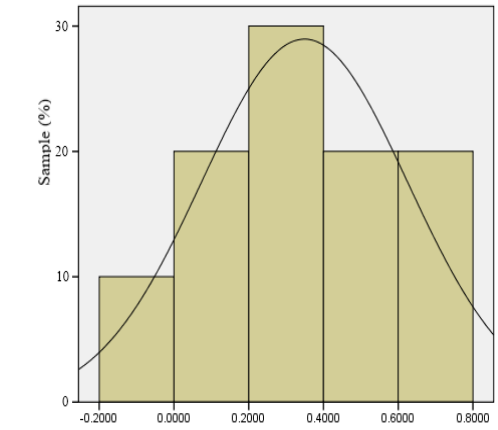
Table 7.1: Characteristics of the sample (study 1)

Age (years)	33.3±9.4
Marital status	
Married	0%
Single, Divorced	100%
University/post diploma	90%
Work hours/week	39.2±1.7
Duration between two questionnaires (minutes)	221.3±37.0
Morning intervals (minutes)	250.3±37.5
Afternoon intervals (minutes)	192.3±41.2
Number of reported positive mood changes	4.9±1.5
Number of reported negative mood changes	5.4±2.0

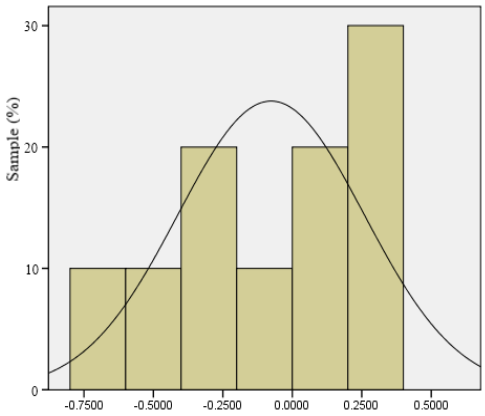
The paper-based questionnaires were administered at 10:00, 14:00 and 18:00 (or with slight deviations when subjects were temporary unable to fill-out the questionnaire) thus dividing working day in two intervals of interest – one between 10:00 and 14:00 and another between 14:00 and 18:00. The amount of speech activity was expressed as the number of minutes in which speech status was identified, divided by the duration of the monitored interval. In total, 122 questionnaires were collected and the the self-reported mood dimensions of PA and NA were analyzed with respect to the amount of speech activity detected in the previous time interval. Overall, 78 such intervals were analyzed, with the duration of 221±37 minutes, in which subjects spent 27.9±12.1% of time (minutes) in speech activity.

7.2.2 Results

Figure 7.1 shows the distribution of Spearman correlation between the amount of speech activity as estimated from accelerometer readings and reported mood changes. The mean correlation between the amount of speech activity and PA and NA scores was 0.34±0.27 (*min*=-0.03, *max*=0.76) and -0.07±0.33 (*min*=-0.62, *max*=0.39) respectively.



a)



b)

Figure 7.1: Distributions of Spearman correlations between an amount of speech activity and a) PA and b) NA

The distribution of the correlations between the amount of detected speech and PA scores were significantly greater than 0 (t=4.009, P<0.005) and not significantly skewed. The distribution related to NA scores was not significantly less than 0 (t=-0.721) and was significantly negatively skewed. On the other hand, the mood score reported at the beginning of monitored interval and the amount of speech activity within that interval showed no significant correlations, 0.153 and 0.225 for PA and NA respectively, indicating that participation in verbal social interactions was not influenced by the initial subjects' mood. This may be due to the fact that working environment typically imposes conversations leaving no options for the one to choose the level of socialization depending on the current state of mood.

The results suggest that the time spent in speech activity (reflecting the participation in verbal social interactions) was positively correlated with changes in reported PA and was not related to the changes in NA scores. Despite being a pilot study which

included 10 subjects, the experimental setting was fully unconstrained yielding results that are consistent with the previous research [117][118][119] which reported positive association between the mood dimension of PA and the amount of speech activity. Such findings indicate the feasibility of using the proposed automatic method of collecting social interaction data for exploring how certain aspects of social interactions affect the psychological response in individuals.

However a positive correlation has been found in previous studies between NA and specific types of social interactions (for instance related arguing or receiving help [115]) rather than with overall amount of social activity. Therefore, examining the context of social interactions can provide more precise insight into the relation between social activity and mood states. In the next study, a set of pleasant social interactions is isolated to examine its impact on reported mood changes.

7.3. Pleasant social interactions and mood changes

Social interaction is typically an integral aspect of work, involving different kinds of conversations – from chats about unimportant matters between colleagues to official meetings, negotiations or interviews. Therefore, social interactions can be perceived from workers' point of view both as a pleasant experience but also as a displeasing one (for example, a small talk with colleagues or an agreement from coworkers versus an imposed meeting or having an argument). This makes the workplace a source of stimulus both for positive and negative emotions.

7.3.1 Monitoring approach

For the purpose of this work, a set of pleasant social interactions was recognized as interactions that occurred during coffee and snack breaks. The term coffee/snack break refers to the place where the interactions occurred during work time; therefore, for the rest of the chapter it is referred to these as breaks. In order to recognize a set of pleasant social interactions and investigate how they affect workers' mood, the approach was based on the location recognition, in particular focusing on break room and balconies. These are the locations where workers (regarding the testbed workplace) typically have breaks during working time and have the opportunity for a relaxed conversation with colleagues. In order to investigate the assumption that

social interactions during breaks are perceived in a positive way, a survey was conducted among 15 colleagues and among participants in the study. They were asked to rate on a 5-point scale (from "not at all" to "very") the statement: "Social interaction during breaks is mostly pleasant for me". The mean score was 3.87 for a randomly chosen sample and 4.00 for participants of this study, varying in answers only from 3 to 5.

The monitoring framework recognizes with a high certainty the location of the subjects, when they are in the break room, meeting room or on the balconies; therefore, as opposed to self-reporting methods, the monitoring system provides much more reliable and precise information about workers' behavior while not intruding in their routines [78]. In addition, speech activity detection was used to confirm the occurrence of a social interaction and to distinguish if a subject made a break alone or in company. In this manner, it is possible to isolate, with a high probability, a part of pleasant social interactions and to assess their influence on the mood.

It should be mentioned that five out of nine recruited subjects were not wearing accelerometer thus the information about speech activity was missing. Therefore, another survey was conducted among the aforementioned groups of workers that were asked to indicate the approximate percent of cases when they are going to a break with someone rather than alone. The possible answers were "<50%", "50-60%", "60-70%", "70-80%", "80-90%" and "90-100%". The results showed that breaks are centered on social interactions considering that one third of each sample reported 80-90%, while 5 participants out of 9 from the first sample and 7 out of 15 from the second sample are very likely (90-100% of cases) to socialize while taking a break. Only one participant (belonging to the randomly chosen sample) reported that he usually takes a break alone.

7.3.2 Experiments

The experiments involved 9 knowledge workers (6 males, 3 females), not connected with this study, for 7 working days within a period of one month (characteristics of the sample is shown in Table 7.2). Subjects were filling out the mood questionnaires in the beginning, in the middle and at the end of working day. There were no significant differences between men and women either in the relevant parameters (such as age, number of working hours or type of the job) or in the measures (such as

a number of reported positive/negative mood changes or an average number of breaks).

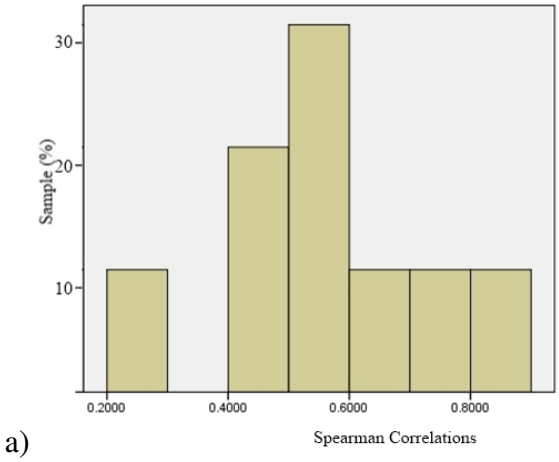
Table 7.2: Characteristics of the sample (study 2)

Age (years)	28.4±2.8
Marital status	
Married	11%
Single, Divorced	89%
University/post diploma	77%
Work hours/week	36.6±4.6
Duration between two questionnaires (minutes)	174.3±49.8
Morning intervals (minutes)	187.4±47.5
Afternoon intervals (minutes)	161.3±48.5
Number of breaks in one interval	1.6±0.7
Number of reported positive mood changes	5.3±1.7
Number of reported negative mood changes	5.7±2.1

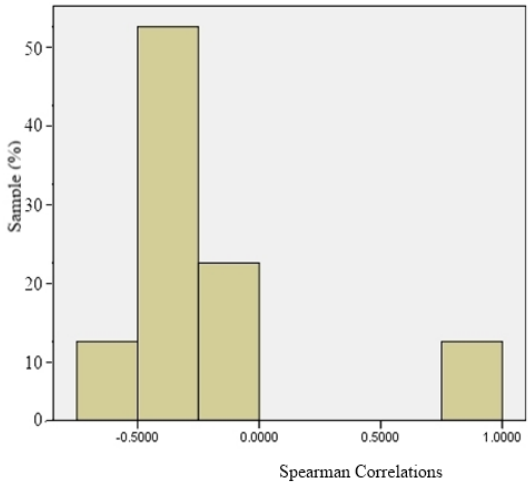
7.3.3 Results

After discarding intervals due to non-completed reports, the data analyzed contained 112 monitored intervals, 54 and 58 intervals of positive and negative mood changes respectively. The overall duration of the recorded data was 339.8 hours, 181.2h in morning intervals (between first and second questionnaire) and 158.6h in afternoon intervals (between second and third questionnaire). The mean score for PA and NA scores was 2.9±0.6 and 2.1±0.7 respectively. The mean within-subject correlation between PA and NA scores was 0.15± 0,09. Self-reported mood change, measured as a difference in scores between two consecutive questionnaires, was analyzed with respect to a number of detected breaks. Spearman correlations were calculated between mood scores and a number of breaks on a within-subject basis. Statistcal analysis was performed using SPSS.

The distributions of Spearman correlations regarding number of breaks and reported mood score changes are shown in Figure 7.2.



a)



b)

Figure 7.2: Distributions of Spearman correlations between number of breaks and a) PA, b) NA

The mean correlation between number of breaks and positive mood changes was 0.57±0.15 (*min*=0.29, *max*=0.83) Figure 7.2a; between number of breaks and negative mood changes was -0.21±0.43 (*min*=-0.66, *max*=0.86) Figure 7.2b. The distribution of Spearman correlations between number of breaks and PA was significantly greater than 0 (t=3.701, P<0.005). However, the distribution of the Spearman correlation between number of breaks and NA was not significantly different from 0 and it was significantly positively skewed. The results of this study indicate that the subjects were more likely to improve their mood when they had breaks, according to reported PA scores. On the other hand, no associations have been found between number of breaks and NA score reports.

Considering the results from the related studies, Vittengl et al. [115] reported positive correlation between PA and fun/active and informational types of social in-

teractions. Therefore, the results reported in this section are consistent with the previously reported findings. The findings indicate the possibility of using the technologically-based approach for extracting relevant parameters for the study of mood changes as an alternative to the use of more cumbersome surveys.

7.4. Impact of individuals on mood changes

The previous two studies demonstrated that relying on the speech detection and localization affords the investigation of parameters that may have an impact on the mood. However, using the fusion of spatial settings recognition and speech activity detection provides a more complete description of social interactions which include the identification of subjects that were engaged in social interactions. This section investigates correlations between mood states and the parameters recognized using the fusion of the two systems.

7.4.1 Experiments and Results

As previously demonstrated, detection of social interaction is based on analyzing spatial parameters between a pair of subjects that carry mobile phones and accelerometer-based speech activity inference. If more than two subjects are involved in the same conversation, the method recognizes other participants by examining information for each pair of individuals involved in the social interaction. Therefore, this allows reconstructing social interaction occurrences among monitored subjects and the exact amount of time that each pair of subjects spent interacting. In this study, it is examined if social interactions between certain subjects induced consistent responses in their mood.

The experimental data used for this study was acquired during the same continuous trial described in Section 5.2 which ran for 7 working days involving four subjects (3 males and 1 female) that share the same office. Similarly to the previous two studies, the subjects were filling out the mood questionnaire three times a day, at 11h, 14h and 17h. The characteristics of the sample are presented in Table 7.3. The data collection resulted in 75±12 hours related to morning intervals and 73±10 hours recorded in afternoon intervals.

Table 7.3: Characteristics of the sample (study 3)

Age (years)	29.2±1.7
Marital status	
Married	25%
Single, Divorced	75%
University/post diploma	100%
Work hours/week	38.0±2.3
Duration between two questionnaires (minutes)	179.0±29.3
Morning intervals (minutes)	184.3±31.1
Afternoon intervals (minutes)	173.7±38.2
Number of reported positive mood changes	4.8±1.0
Number of reported negative mood changes	5.2±1.9

The mood scores were analyzed for each interval with respect to the amount of social interactions between each pair of subjects. Despite the existing manual annotations, social encounters were reconstructed relying on the spatial settings and speech activity analysis including also the data that was not annotated due to the observers' lack of presence (21% of data). Amount of social interactions between each pair of subjects was expressed as the percentage of time that they spent in conversations with respect to the duration of the interval (regardless if conversations included exclusively these two subjects or other individuals as well). For each pair of subject there were analyzed 12.3±1.0 such intervals, with the duration of 179.0±29.3 minutes. Note that four subjects make six pairs and that each monitored working day result in two intervals, denoted as morning or afternoon interval. The mood changes were calculated as a difference in scores in the beginning and at the end of each interval across PA and NA adjectives. Within-subjects correlation was calculated for each participant with respect to social interactions with other three monitored subjects. However, no statistically significant correlation was found either for PA or NA reports.

The results suggest that the overall amount of time that one subject spent talking to a certain individual does not affect her/his mood but the mood changes depend more on the context of social interactions, referring to the current literature and the results provided in Section 7.3). In other words, the mood changes are more related to *how* than to *with whom* subjects pre-dominantly socially interacted.

However, the experiments yielded an interesting result regarding the relation between social interactions and the absolute mood scores measured in the beginning

of the monitored interval. When the amount of time that one subject spent talking to other monitored subjects from the office (that were also monitored) was subtracted from the total amount of time he/she spent speaking (estimated using accelerometer), the remaining portion refers to conversations with all other individuals but the office colleagues. It was discovered that the amount of this portion of social interaction and the absolute mood states reported at the beginning of monitored intervals showed significant correlations for three out of four subjects regarding PA adjectives (Table 7.4).

Table 7.4: Correlation between self-reported mood and the amount of social interactions outside of the office

	Positive Mood adjectives	Negative mood adjectives
Subject 1	0.380*	0.135
Subject 2	0.299*	0.015
Subject 3	0.351*	-0.054
Subject 4	0.171	0.120

$*P < 0.05$

The explanation for such result may be that the conversations in the office are mostly imposed by colleagues while the level of socialization outside of the office depended on the subjects' current mood (excluding work related meetings).

7.5. Summary

Social activity is linked to various psycho-physical health outcomes and to the overall wellbeing. In this regard, monitoring and assessing socialization patterns of individuals become an important aspect, typically addressed relying on self-reporting methods. In addition to the common drawbacks of memory dependence and a high user effort, using surveys creates a unique set of concerns when studying psychological aspects in individuals, considering the fact that self-reports are strongly influenced by the mood of subject.

The current literature reports several studies that examined how the social activity impacts the mood during the day, while none relied on the automated methods for collecting data.

This chapter described the use of the proposed solution to investigate the correlation between various parameters of social activity and the mood changes in office workers. The results suggest the positive correlation between the amount of social activity and the mood dimension of PA, while no evidenced correlations were found related to NA. Despite being a pilot study, the experimental setting was fully unconstrained yielding results that are consistent with the previous research. Therefore, this study demonstrated the possibility of using the proposed sensor-based method for monitoring aspects of social interactions that are relevant for similar investigations in the domain of social psychology.

Chapter 8

8. Conclusions

The rapid development of technology has significantly contributed to the improvement of a myriad of scientific disciplines, thus enabling their accelerated advancements. Due to the possibility of automatic monitoring of various aspects related to social behavior, social sciences are now at a critical point in their evolution [5]. Technology provides ample opportunities for acquisition and processing of a variety of information, however the challenge remains for the researchers as to how to use these new instruments to conduct a study which approximates the real-life situations. In this regard, the two main issues for automatic social interaction data collection include privacy respecting and obtrusiveness [15] – aspects which directly affect the natural behavior of subjects. Invasive sensors typically provide the output which is easier to process (and vice versa), thus the method for monitoring social interactions reflects the trade-off between the quality of extracted data and the experimental conditions. Most approaches in this domain have utilized video and/or audio systems to sense evidences of an ongoing face-to-face social interaction – visual evidences such as the posture and proximity of subjects, and/or the auditory ones which include speech activity and laugher. However, capturing audio/video data can be perceived as privacy intrusive; since the video systems require direct line of sight between cameras and subjects, they impose yet another drawback which is the restriction of individuals' freedom of movement during experiments.

The work in this thesis demonstrated the feasibility of monitoring co-located social interactions in a continuous and mobile manner which does not utilize visual or auditory data. The proposed approach relies on the sources which do not raise privacy concerns in subjects and do not interfere with their daily routines. Nevertheless, the obtained data can be interpreted in order to infer spatial settings between subjects and their speech activity with a sufficiently high accuracy for a reliable social interaction

data collection. The specific contributions of this thesis are summarized in the following.

This thesis presented the method of detecting spatial settings between subjects, described through parameters of interpersonal distances and relative body orientation, relying solely on mobile phone sensing capabilities. The previous work suggested the possibility of inferring social interactions based on interpersonal distances and relative body orientations detected using camera system; however the work in this thesis demonstrated that there can be a trade-off between accuracy in recognizing the two spatial parameters and providing a mobile phone-based solution which does not use video/audio data. The distance estimation required the adaptation of RSSI fingerprinting algorithm for estimating the distance between two mobile phones, resulting in the median accuracy of 0.5 m, which represents an improvement in comparison to the state of the art systems.

In addition, the approach of identifying speech activity status using an off-the-shelf accelerometer that can identify the vibrations of vocal chords was proposed as an alternative method to microphone-based speech detection, thus preserving the privacy of subjects.

This work evaluated the performance of face-to-face interaction detection which is based on the obtained information on spatial settings and speech activity status. The experiments evidenced the highest reliability in the inference of social interactions when having the knowledge of both spatial settings and speech activity status of individuals, while also indicating several situations in which only spatial settings sufficed.

Furthermore, the proposed solution provisions an automatic classification of the type of social context. The analysis showed the high predictive power of standard deviation of relative body orientation as a classification feature both for inferring the occurrence of social interactions and for recognizing the social context.

The feasibility of using the proposed system was further demonstrated in collecting the data relevant for investigating psychological effects of social activities.

The proposed concept of monitoring social interactions preserves the privacy of monitored subjects since it does not capture video or audio data which is typically considered to be sensitive. Despite this, any kind of data on human behavior can also

contain private information and therefore raise privacy concerns. However, considering the computational capabilities of the current mobile phones, the work in this thesis allows the development of a monitoring platform which would store all the inferences locally on a mobile phone, enabling users to select the data that they wish to be shared for the experiments.

The obtrusiveness of the approach is minimized by using a mobile phone, which is a widely adopted device, and an accelerometer with small dimensions, in comparison to typical devices dedicated to sensing social interactions. It is also reasonable to expect that accelerometer devices will further shrink in size thus making them less obtrusive to be attached onto the body.

8.1. Future work

The work in this thesis provides the possibility to extract a unique set of nonverbal cues in a mobile way, related to spatial settings and vocal behavior. As a first step in exploring social signals using the proposed system, it was demonstrated that some of these cues exhibit the high predictive power in classifying the social context and in inferring of social interactions. The proposed mobile monitoring platform opens up a number of possible directions for exploring aspects of social interactions that were not approachable using traditional techniques or the sensor-based systems for continuous monitoring.

In addition, a set of extracted nonverbal cues can be extended without involving additional sensors. The evaluation of the accelerometer-based approach indicated that the resolution of 10 s was not sufficient for exploring a range of features related to nonverbal vocal behavior, including turn-taking patterns, successful and unsuccessful interruptions, and the amount of silence. This may be addressed using a different type of the accelerometer, thus potentially identifying speech status on a fine granularity, which is the issue also experienced with microphone-based approaches. The same problem has constrained researchers to manually annotate speech activity in audio recordings in order to analyze specific vocal behavior features [22][92]. Furthermore, the accelerometer position in the proposed design allows for the extraction of the for-

ward leaning level, a nonverbal cue typically explored with respect to the attitude of subjects in a face-to-face conversation.

Automatic recognition of social activity does not only mean improved data collection for social studies, but it promises a myriad of possible applications which can provide benefits to the users. The analysis of information generated from mobile sensors to infer social interactions as proposed has a potential applications in mobile computing, including content prediction and network resource assignment. These applications are based on social connections between people and higher probability of socially connected people to consume similar content. In addition, since the participation in social activities reflects the status of wellbeing, the proposed solution can form part of a persuasive feedback application for encouraging healthier lifestyle. Stimulating healthier lifestyle through such applications is based on the concept of providing people with self-monitoring tools, which increase the awareness of their daily routines and consequently their wellbeing.

8.2. Final remarks

Considering the trends of technological advancements, it is reasonable to expect that the technologies used in this thesis will at some point in future become obsolete. However, the work in this thesis provides a concept for monitoring social activity while avoiding the capturing of visual and auditory information. The drawbacks of the current sensor-based methods may be the rationale behind why self-reports are still prevalent for collecting social interaction data. Neither the current systems nor the approach presented in this thesis is still a suitable replacement of the gold standard surveys for a number of studies. However, addressing shortcomings of the current sensor-based collecting methods for monitoring social interactions and decreasing negative effects of the observation will lead towards their wider acceptance and this thesis is envisioned to be a step towards this goal.

Bibliography

[1] H. Triplett, "The dynamogenic factors in pacemaking and competition," *American Journal of Psychology*, vol. 9, pp. 507–533, 1898.

[2] F. J. Roethlisberger and W. J. Dickson, *Management and the worker*, vol. 21, no. 202/203. Harvard University Press, 1939, pp. pp.306–308.

[3] M. B. Parten, "Social participation among pre-school children," *The Journal of Abnormal and Social Psychology*, vol. 27, no. 3, pp. 243–269, 1932.

[4] K. Lewin, R. Lippitt, and R. K. White, "Patterns of aggressive behavior in experimentally created social climates," *Journal of Social Psychology*, vol. 10, no. 2, pp. 271–299, 1939.

[5] N. N. Eagle, "Machine perception and learning of complex social systems," Massachusetts Institute of Technology, 2005.

[6] M. Rabbi, T. Choundhury, S. Ali, and E. Berke, "Passive and in-situ assessment of mental and physical wellbeing using mobile sensors," in *13th International Conference on Ubiquitous Computing (UbiComp'11)*, 2011.

[7] H. R. Bernard, P. Killworth, D. Kronenfeld, and L. Sailer, "The Problem of Informant Accuracy: The Validity of Retrospective Data," *Annual review of anthropology*, vol. 13, no. 1, pp. 495–517, 1984.

[8] T. Choudhury, "Sensing and modeling human networks," Massachusetts Institute of Technology, 2004.

[9] A. A. Salah, M. Pantic, and A. Vinciarelli, "Recent developments in social signal processing," in *IEEE International Conference on Systems, Man and Cybernetics*, 2011, pp. 380–385.

[10] D. Olguin Olguin and A. S. Pentland, "Social sensors for automatic data collection," *14th Americas Conference on Information Systems*, pp. 1–10, 2008.

[11] M. Buchanan, "The science of subtle signals," *Strategy+Business*, pp. 48:68–77, 2007.

[12] G. Groh, A. Lehmann, J. Reimers, M. R. Frieß, and L. Schwarz, "Detecting social situations from interaction geometry," in *IEEE International Conference on Social*

Computing/IEEE International Conference on Privacy, Security, Risk and Trust, 2010.

[13] D. Wyatt, T. Choudhury, and J. Bilmes, "Inferring colocation and conversation networks from privacy-sensitive audio with implications for computational social science," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 1, 2011.

[14] "Fitbit." [Online]. Available: http://www.fitbit.com/. [Accessed: 10-Mar-2012].

[15] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: survey of an emerging domain," *Image and Vision Computing*, vol. 27, no. 12, pp. 1743–1759, Nov. 2009.

[16] Y. Lee and O. Kwon, "Information privacy concern in context-aware personalized services : results of a delphi study," *Journal of Information Systems*, vol. 20, no. 2, 2010.

[17] D. H. Nguyen, A. Kobsa, and G. R. Hayes, "An empirical investigation of concerns of everyday tracking and recording technologies," *Proceedings of the 10th international conference on Ubiquitous computing - UbiComp '08*, p. 182, 2008.

[18] J. A. Scheinkman, "Social interactions," in *The New Palgrave Dictionary of Economics*, 2nd ed., S. Durlauf and L. Blume, Eds. Palgrave Macmillan, 2008.

[19] E. Hall, *The hidden dimnesion*. New York: Double Day Anchor Books, 1966.

[20] A. Madan, M. Cebrian, D. Lazer, and A. Pentland, "Social sensing for epidemiological behavior change," *12th ACM international conference on Ubiquitous computing*, 2010.

[21] G. W. Allport, "The historical background of social psychology," in *Psychology applied to work*, G. Lindzey and E. Aronson, Eds. Lawrence Erlbaum Associates, Inc., 1985, pp. 1–46.

[22] D. B. Jayagopi, "Computational modeling of face-to-face social interaction using nonverbal behavioral cues," Lausanne, EPFL, 2011.

[23] M. L. Knapp and J. A. Hall, *Nonverbal communication in human interaction*. Holt, Rinehart and Winston, 1972, p. 512.

[24] D. Gatica-Perez, "Automatic nonverbal analysis of social interaction in small groups: A review," *Image and Vision Computing*, vol. 27, no. 12, pp. 1775–1787, Nov. 2009.

[25] N. Eagle and A. (Sandy) Pentland, "Reality mining: sensing complex social systems," *Personal and Ubiquitous Computing*, vol. 10, no. 4, pp. 255–268, Nov. 2005.

[26] K. Greene, "10 emerging tehcnologies 2008," *MIT Technology Review*, 2008.

[27] A. Vinciarelli, M. Pantic, D. Heylen, C. Pelachaud, I. Poggi, F. D. Errico, and M. Schr, "Bridging the gap between social animal and unsocial machine : a survey of social signal processing," *EEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 69–87, 2012.

[28] C. BenAbdelkader, "Statistical estimation of human anthropometry from a single uncalibrated image," *Computational Forensics, Springer*, pp. 1–17, 2008.

[29] Y. Yacoob, "Detection and analysis of hair," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 7, pp. 1164–1169, 2006.

[30] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance," in *In European Conference on Computer Vision*, 2006.

[31] H. Gunes and M. Piccardi, "Assessing facial beauty through proportion analysis by image processing and supervised learning," *International Journal of Human-Computer Studies*, vol. 64, no. 12, pp. 1184–1199, 2006.

[32] H. Gunes, M. Piccardi, and T. Jan, "Comparative beauty classification for pre-surgery planning," in *IEEE SMC 2004 the International Conference on Systems*, 2004, vol. 3, pp. 2168–2174.

[33] A. Sepehri, Y. Yacoob, and L. S. Davis, "Employing the hand as an interface device," *Journal of Multimedia*, vol. 1, no. 7, pp. 18–29, 2006.

[34] W. W. Kong and S. Ranganath, "Automatic hand trajectory segmentation and phoneme transcription for sign language," *2008 8th IEEE International Conference on Automatic Face Gesture Recognition*, pp. 1–6, 2008.

[35] M. Cristani, A. Pesarin, and A. Vinciarelli, "Look at who's talking: voice activity detection by automated gesture analysis," in *Workshop on Interactive Human Behavior Analysis in Open or Public Spaces*, 2011.

[36] R. Poppe, "Vision-based human motion analysis: an overview," *Computer Vision and Image Understanding*, vol. 108, pp. 4–18, 2007.

[37] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding*, vol. 104, no. 2–3, pp. 90–126, 2006.

[38] G. Paggetti, A. Vinciarelli, I. Italiano, and G. It, "Towards computational proxemics : inferring social relations from interpersonal distances," in *IEEE International Conference on Social Computing*, 2011.

[39] M. Cristani and V. Murino, "Socially intelligent surveillance and monitoring: analysing social dimensions of physical space," in *International Workshop on Socially Intelligent Surveillance and Monitoring*, 2010, pp. 51–58.

[40] “Sociometric Solutions.” [Online]. Available: http://www.sociometricsolutions.com/. [Accessed: 25-Mar-2012].

[41] K. Fischbach, P. a. Gloor, and D. Schoder, “Analysis of Informal Communication Networks – A Case Study,” *Business & Information Systems Engineering*, vol. 1, no. 2, pp. 140–149, Dec. 2008.

[42] D. Olguin Olguin, B. N. Waber, T. Kim, A. Mohan, K. Ara, and A. Pentland, “Sensible organizations: technology and methodology for automatically measuring organizational behavior.,” *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society*, vol. 39, no. 1, pp. 43–55, Feb. 2009.

[43] D. Olguın and P. Gloor, “Capturing individual and group behavior with wearable sensors,” *AAAI Spring Symposium on Human Behavior Modeling. Stanford, CA.*, 2009.

[44] C. Cattuto, W. V. den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, and A. Vespignani, “Dynamics of person-to-person interactions from distributed RFID sensor networks,” *PLoS ONE 5(7): e11596. doi:10.1371/journal.pone.0011596*, 2010.

[45] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaggiotto, W. Van Den Broeck, C. Régis, B. Lina, and P. Vanhems, “High-resolution measurements of face-to-face contact patterns in a primary school,” *PLoS ONE*, vol. 6, no. 8, p. 13, 2011.

[46] A. Barrat, C. Cattuto, V. Colizza, L. Isella, A. E. Tozzi, and W. V. D. Broeck, “Wearable sensor networks for measuring face-to-face contact patterns in healthcare settings,” in *3rd International ICST Conference on Electronic Healthcare for the 21st century*, 2010, pp. 1–4.

[47] A. Barrat, C. Cattuto, M. Szomszor, and W. V. D. Broeck, “Social dynamics in conferences: analysis of data from the live social semantics application,” in *Proceedings of the 9th International Semantic Web Conference (ISWC 2010)*, 2010, pp. 17–33.

[48] “The Electronically Activated Recorder (EAR).” [Online]. Available: http://dingo.sbs.arizona.edu/~mehl/EAR.htm. [Accessed: 15-Mar-2012].

[49] M. R. Mehl and M. L. Robbins, *Naturalistic observation sampling: The Electronically Activated Recorder (EAR): Handbook of research methods for studying daily life*. New York, NY: Guilford Press., 2012.

[50] N. Ramirez-Esparza, M. R. Mehl, J. Alvarez Bermudez, and J. W. Pennebaker, “Are Mexicans more or less sociable than Americans? Insights from a naturalistic observation study,” *Journal of Research in Personality*, vol. 43, no. 1, pp. 1–7, 2009.

[51] M. R. Mehl, S. Vazire, N. Ramirez-Esparza, R. B. Slatcher, and J. W. Pennebaker, "Are women really more talkative than men?," *Science*, vol. 317, no. 82, 2007.

[52] M. R. Mehl, S. D. Gosling, and J. W. Pennebaker, "Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life," *Journal of Personality and Social Psychology*, vol. 90, pp. 862–877, 2006.

[53] M. R. Mehl, S. Vazire, S. E. Holleran, and C. S. Clark, "Eavesdropping on happiness: Well-being is related to having less small talk and more substantive conversations," *Psychological Science*, vol. 21, pp. 539–541, 2010.

[54] N. S. Holtzman, S. Vazire, and M. R. Mehl, "Sounds like a narcissist: Behavioral manifestations of narcissism in everyday life," *Journal of Research in Personality*, vol. 44, pp. 478–484, 2010.

[55] S. E. Holleran, J. Whitehead, T. Schmader, and M. R. Mehl, "Talking shop and shooting the breeze: A study of workplace conversations and job disengagement among STEM faculty," *Social Psychological and Personality Science*, vol. 2, pp. 65–71, 2011.

[56] R. Borovoy, F. Martin, S. Vemuri, M. Resnick, B. Silverman, and C. Hancock, "Meme tags and community mirrors: moving from conferences to collaboration," in *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, 1998, vol. 98, pp. 159–168.

[57] "Make your meetings matter: SpotMe." [Online]. Available: http://www.spotme.com/. [Accessed: 29-Mar-2012].

[58] "Alliance Tech." [Online]. Available: http://www.alliancetech.com/. [Accessed: 12-Mar-2012].

[59] L. Holmquist, "Supporting Group Collaboration with IPAD: s-Inter-Personal Awareness Devices," *Proc. Workshop on Handheld CSCW, ACM CSCW*, pp. 105–124, 1998.

[60] N. Eagle and A. (Sandy) Pentland, "Reality mining: sensing complex social systems," *Personal and Ubiquitous Computing*, vol. 10, no. 4, pp. 255–268, Nov. 2005.

[61] C. A. Hidalgo and C. Rodriguez-Sickert, "The dynamics of a mobile phone network," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 12, pp. 3017–3024, 2007.

[62] M. Raento, A. Oulasvirta, R. Petit, and H. Toivonen, "ContextPhone: A prototyping platform for context-aware mobile applications," *IEEE Pervasive Computing*, vol. 4, no. 2, pp. 51–59, Apr. 2005.

[63] S. Mardenfeld, D. Boston, S. J. Pan, Q. Jones, A. Iamntichi, and C. Borcea, "GDC: Group Discovery using Co-location traces," *2010 IEEE Second International Conference on Social Computing*, pp. 641–648, 2010.

[64] T. M. T. Do and D. Gatica-Perez, "GroupUs: Smartphone proximity data and human interaction type mining," in *5th annual International Symposium on Wearable Computers*, 2011, no. 2.

[65] N. Banerjee, S. Agarwal, P. Bahl, R. Chandra, A. Wolman, and M. Corner, "Virtual compass: relative positioning to sense mobile social interactions," *Pervasive Computing*, pp. 1–21, 2010.

[66] R. Sommer, *Personal Space*. Prentice-Hall, Inc., 1959.

[67] E. GAINES, "Communication and the semiotics of space," *Journal of Creative Communications*, vol. 1, no. 2, 2006.

[68] M. Hazas, C. Kray, H. Gellersen, H. Agbota, G. Kortuem, and A. Krohn, "A relative positioning system for co-located mobile devices," *Proceedings of the 3rd international conference on Mobile systems, applications, and services - MobiSys '05*, p. 177, 2005.

[69] C. Peng, G. Shen, Y. Zhang, and Y. Li, "Beepbeep: a high accuracy acoustic ranging system using cots mobile devices," *Proceeding SenSys '07 Proceedings of the 5th international conference on Embedded networked sensor systems*, 2007.

[70] N. Eagle, A. S. Pentland, and D. Lazer, "Inferring friendship network structure by using mobile phone data.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 36, pp. 15274–8, Sep. 2009.

[71] J. Krumm and K. Hinckley, "The NearMe Wireless Proximity Server," *In Proceedings of International Conference on Ubiquitous Computing*, pp. 283–300, 2004.

[72] P. Bhagwat, B. Raman, and D. Sanghi, "Turning 802.11 inside-out," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 1, p. 33, Jan. 2004.

[73] A. Popleteev, "Indoor positioning using FM radio signals," DISI - University of Trento, 2011.

[74] K. Kaemarungsi, "Distribution of WLAN received signal strength indication for indoor location determination," *2006 1st International Symposium on Wireless Pervasive Computing*, pp. 1–6, 2006.

[75] B. Ferris, D. Hahnel, and D. Fox, "Gaussian Processes for Signal Strength-Based Location Estimation," *Robotics: Science and Systems II*, 2006.

[76] W. North, *Proxemics: The Semiotics of Space*. Indiana University Press, 1995.

[77] I. Carreras, A. Matic, P. Saar, and V. Osman, “Comm2Sense: Detecting Proximity Through Smartphones,” in *PerMoby Workshop, in IEEE PerCom Conference*, 2012.

[78] A. Matic, A. Papliatseyeu, V. Osmani, and O. Mayora-Ibarra, “Tuning to your position: FM radio based indoor localization with spontaneous recalibration,” *2010 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 153–161, Mar. 2010.

[79] A. Mehrabian, “Some referents and measures of nonverbal behavlor,” *Behavior Research Methods and Instrumentation*, vol. 1, no. 6, pp. 203–207, 1969.

[80] Y. Shi, Y. Shi, and J. Liu, “A rotation based method for detecting on-body positions of mobile devices,” in *Proceedings of the 13th International Conference on Ubiquitous Computing*, 2011.

[81] R. Rao and T. Chen, “Cross-modal prediction in audio-visual communication,” in *IEEE international Conference on Acoustics, Speech, and Signal Processing. ICASSP-96.*, 1996, pp. 2056–2059.

[82] J. Sundberg, “Chest wall vibrations in singers.,” *Journal Of Speech And Hearing Research*, vol. 26, no. 3, pp. 329–340, 1983.

[83] T. H. Falk, J. Chan, P. Duez, G. Teachman, and T. Chau, “Augmentative communication based on realtime vocal cord vibration detection.,” *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 18, no. 2, pp. 159–63, Apr. 2010.

[84] T. F. Quatieri, F. Ieee, K. Brady, D. Messing, J. P. Campbell, W. M. Campbell, M. S. Brandstein, S. M. Ieee, C. J. Weinstein, J. D. Tardelli, and P. D. Gatewood, “Exploiting Nonacoustic Sensors for Speech Encoding,” *Language*, vol. 14, no. 2, pp. 533–544, 2006.

[85] “Medicine Net.” [Online]. Available: http://www.medterms.com/script/ main/ art.asp?articlekey=6224. [Accessed: 15-Nov-2011].

[86] I. Titze, “Physiologic and acoustic differences between male and female,” *J. Acoust. Soc. Am*, pp. 1699–1707, 1989.

[87] M. J. Mathie, A. C. F. Coster, N. H. Lovell, and B. G. Celler, “Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement,” *Physiological Measurement*, vol. 25, no. 2, pp. R1–R20, Apr. 2004.

[88] “Shimmer - Wireless Sensor Platform for Wearable Applications.” [Online]. Available: http://www.shimmer-research.com/p/products/sensor-units-and-modules/wireless-ecg-sensor . [Accessed: 15-Nov-2011].

[89] S. M. Tan, “Chapter 9 The Discrete Fourier transform,” in *Linear Systems*, The University of Auckland, pp. 1–8.

[90] B. Widrow, J. R. G. Jr, and J. M. McCool, “Adaptive noise cancelling: Principles and applications,” *Proceedings of the IEEE*, vol. 63, no. 12, pp. 105–112, 1975.

[91] G. Groh and C. Fuchs, “Combining evidence for social situation detection,” *Proc. IEEE Socialcom2011*, 2011.

[92] G. Groh, A. Lehmann, and M. D. Souza, “Mobile Detection of Social Situations with Turn Taking Patterns,” in *WAC2011*, 2011.

[93] C. Savage, *Fifth Generation Management, Second Edition: Dynamic Teaming, Virtual Enterprising and Knowledge Networking*. Butterworth-Heinemann, 1996.

[94] U. M. Apte and H. K. Nath, “Size, structure and growth of the U.S. information economy.,” in *Managing in the in- formation economy. Springer, Heidelberg, pp 1–28*, Springer, Heidelberg, 2007, pp. 1–28.

[95] M. Mcdermott, “Knowledge Workers: You can gauge their effectiveness,” *Leadership Excellence*, vol. 22, 2005.

[96] R. Cross, A. Parker, and L. Sasson, *Networks in the knowledge economy*. Oxford University Press, Oxford, 2003.

[97] S. Aral, E. Brynjolfsson, and M. V. Alstyne, “Information, technology and information worker productivity: task level evidence,” in *Proceedings of the 27th annual international conference on information systems*, 2006.

[98] R. S. Fish, R. W. Root, and B. L. Chalfonte, “Informal communication in organizations: Form, function, and technology,” in *Human reactions to technology: Claremont symposium on applied social psychology*, 1990.

[99] R. Aalbers, O. Koppius, and W. Dolfsma, “On and off the beaten path: Transferring knowledge through formal and informal networks,” *CIRCLE Electronic Working Paper Series*, 2006.

[100] D. Krackhardt and J. R. Hanson, “Informal networks: the company behind the charts,” *Harvard Business Review*, vol. 74, no. 4, pp. 104–111, 1993.

[101] M. ME, “What makes information workers productive,” *MIT Sloan Management Review*, vol. 49, no. 2, pp. 16–17, 2008.

[102] R. A. Hannemann and M. Riddle, *Introduction to social network methods*. University of California, Riverside, 2005.

[103] B. P and H. DA, “Ties, leaders, and time in teams – strong inference about net-work structure’s effects on team viability and performance,” *Academy of Management Journal*, vol. 49, no. 1, pp. 49–68, 2006.

[104] N. Eagle, A. S. Pentland, and D. Lazer, “Inferring social network structure using mobile phone data,” *Proceedings of the National Academy of Sciences (PNAS)*, vol. 106, no. 6, pp. 15274–15278, 2009.

[105] S. Whittaker, D. Frohlich, and O. Daly-Jones, “Informal workplace communication: What is it like and how might we support it?,” in *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence*, 1994, pp. 131–137.

[106] D. N. E. and J. K. Burgoon, “Perceptions of power and interactional dominance in inter- personal relationships.,” *Journal of Social and Personal Relationships*, vol. 22, no. 2, pp. 207–233, 2005.

[107] M. Schmid Mast, “Dominance as expressed and inferred through speaking time: a meta- analysis,” *Human Communication Research*, vol. 28, no. 3, pp. 420–450, 2002.

[108] L. (1989). Brody, C. and Smith-Lovin, “Interruptions in group discussions: The effects of gender and group composition,” *American Sociological Review*, vol. 54, no. 3, pp. 424–435, 1989.

[109] A. Bhattacharyya, “On a measure of divergence between two statistical populations defined by their probability distributions,” *Bulletin of the Calcutta Mathematical Society*, vol. 35, pp. 99–109, 1943.

[110] J. S. House, K. R. Landis, and D. Umberson, “Social relationships and health.,” *Science (New York, N.Y.)*, vol. 241, no. 4865, pp. 540–5, Jul. 1988.

[111] V. Isaac, R. Stewart, S. Artero, M.-L. Ancelin, and K. Ritchie, “Social activity and improvement in depressive symptoms in older people: a prospective community cohort study.,” *The American journal of geriatric psychiatry official journal of the American Association for Geriatric Psychiatry*, vol. 17, no. 8, pp. 688–696, 2009.

[112] H. B. Bosworth, J. C. Hays, L. K. George, and D. C. Steffens, “Psychosocial and clinical predictors of unipolar depression outcome in older adults.,” *International Journal of Geriatric Psychiatry*, vol. 17, no. 3, pp. 238–246, 2002.

[113] N. D. Lane, T. Choudhury, A. Campbell, M. Mohammod, M. Lin, X. Yang, A. Doryab, H. Lu, S. Ali, and E. Berk, “BeWell: A smartphone application to monitor, model and promote wellbeing,” in *5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth2011)*, 2011.

[114] J. J. a Denissen, L. Butalid, L. Penke, and M. a G. van Aken, “The effects of weather on daily mood: a multilevel approach.,” *Emotion (Washington, D.C.)*, vol. 8, no. 5, pp. 662–7, Oct. 2008.

[115] J. R. Vittengl and C.S.Holt, “A Time-series diary study of mood and social interaction,” *Motivation and Emotion*, vol. 22, no. 3, pp. 255–275, 1998.

[116] P.R. Robbins and R. H. Tanck, “A study of diurnal patterns of depressed mood,” *Motivation and Emotion*, vol. 11, no. 1, pp. 37–49, 1987.

[117] D. S. Berry and J. S. Hansen, “Positive affect, negative affect, and social interaction.,” *Journal of Personality and Social Psychology*, vol. 71, no. 4, pp. 796–809, 1996.

[118] L. A. Clark and D. Watson, “Mood and the mundane: relations between daily life events and self-reported mood.,” *Journal of Personality and Social Psychology*, vol. 54, no. 2, pp. 296–308, 1988.

[119] D. Watson, L. A. Clark, C. W. McIntyre, and S. Hamaker, “Affect, personality, and social activity.,” *Journal of Personality and Social Psychology*, vol. 63, no. 6, pp. 1011–1025, 1992.

[120] L. A. Clark, D. Watson, and J. Leeka, “Diurnal variation in the possitive affect,” *Motivation and Emotion*, vol. 13, no. 3, pp. 205–134, 1999.

[121] A. Adan and J. Guàrdia, “Circadian variations of self-reported activation: a multidimensional approach.,” *Chronobiologia*, vol. 20, no. 3–4, pp. 233–244.

[122] A. C. Volkers, J. H. M. Tulen, and W. W. V. D. Broek, “Relationships between sleep quality and diurnal variations in mood in healthy subjects,” 1998. [Online]. Available: http://www.nswo.nl/userfiles/files/publications/jaarboek-1998/volkers.pdf.

[123] J. M. Smyth, “Ecological Momentary Assessment Research in Behavioral medicine,” *Journal of Happiness Studies*, vol. 4, no. 1, pp. 35–52, 2003.

[124] C. D. Spielberger, “Profile of Mood States.,” *Professional Psychology*, vol. 3, no. 4, pp. 387–388, 1972.

Appendix A – Relevant publications

1. Aleksandar Matic, Venet Osmani, Oscar Mayora, "Analysis of Social Interactions through Mobile Phones", Journal of Mobile Networks and Applications (MONET), DOI: 10.1007/s11036-012-0400-4, 2012.

2. Aleksandar Matic, Venet Osmani, Oscar Mayora, "Automatic Sensing of Speech Activity and Correlation with Mood Changes", Book Chapter in Pervasive Mobile Sensing and Computing for HealthCare, S. C. Mukhopadhyay (Eds.), Springer, 2012.

3. Aleksandar Matic, Venet Osmani, Alban Maxhuni, Oscar Mayora, "Multi-Modal Mobile Sensing of Social Interactions", In Proceedings of 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), San Diego, United States, 2012.

4. Aleksandar Matic, Venet Osmani, Oscar Mayora, "Speech activity detection using accelerometer", In Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE EMBS), Aug 28 – Sep 1, San Diego, California, USA, 2012.

5. Aleksandar Matic, Venet Osmani, Andrei Popleteev, Oscar Mayora, "Smart Phone Sensing to Examine Effects of Social Interactions and Non-Sedentary Work Style on Mood Changes". In Proceedings of the 7th International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT '11), Karlsruhe, Germany, 2011.

6. Venet Osmani, Aleksandar Matic, Iacopo Carreras, "Detection of Social Interactions through Mobile Phones", 6th International Workshop on Ubiquitous Health and Wellness (UbiHealth), 2012.

7. Piret Saar, Aleksandar Matic, Iacopo Carreras, Venet Osmani, “Proximity Detection via Smart-phones”. 4th ICST International Conference on eHealth, Malaga, Spain, 2011

8. Iacopo Carreras, Aleksandar Matic, Piret Saar, Venet Osmani, "Comm2Sense: Detecting Proximity Through Smart-phones". International Workshop on the Impact of Human Mobility in Pervasive Systems and Applications (PerMoby 2012), in IEEE PerCom Conference, 2012.

9. Alban Maxhuni, Aleksandar Matic, Venet Osmani, Oscar Mayora. "Correlation between self-reported mood states and objectively measured social interactions at work: A Pilot Study". In Proceedings of Pervasive Health MindCare Workshop, Ireland, May 2011

10. Aleksandar Matic, Andrei Papliatseyeu, Venet Osmani, Oscar Mayora: “FM Radio for Indoor Localization with Spontaneous Recalibration”, Journal of Pervasive and Mobile Computing (Elsevier), Vol.6, Issue 6, pp. 642-656, 2010.

11. Aleksandar Matic, Andrei Papliatseyeu, Silvia Gabrielli, Venet Osmani, Oscar Mayora-Ibarra, "Happy or Moody? Why so? Monitoring Daily Routines at Work and Inferring Their Influence on Mood". 5th UbiHealth Workshop in conjunction with UBICOMP 2010 Conference, Copenhagen, Denmark, 2010

.